

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 15:58:31 ; Search time 16 Seconds
(without alignments)
1280.550 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068

Sequence: 1 MKKFLIIAAVAFVAVSADPI.....VRKENTKVLAPAFKRELEKN 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR 78:**

1: pir1:**
2: pir2:**
3: pir3:**
4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	10.2	568	2 T49362	hypothetical prote
2	97.5	9.1	776	2 E90591	lipoprotein [impor
3	93.5	8.8	449	2 F84241	hypothetical prote
4	93.5	8.8	628	2 G75154	tungsten-containin
5	92.5	8.7	373	2 T36830	probable phenylala
6	92.5	8.7	697	2 B86239	protein T10024.19
7	91.5	8.6	979	2 E90601	hypothetical prote
8	91.5	8.6	1393	2 E72122	RNA polymerase bet
9	91.5	8.6	1393	2 B86501	RNA polymerase bet
10	91.5	8.6	1397	2 B81548	DNA-directed RNA p
11	90.5	8.5	547	2 F82783	60kDa chaperonin x
12	90.5	8.5	681	2 T15590	hypothetical prote
13	90	8.4	1558	2 B71503	RESA-H3 antigen pF
14	88.5	8.3	459	2 A46372	immunophilin FKBP5
15	88.5	8.3	2137	2 T05244	hypothetical prote
16	88	8.2	436	2 B95199	phosphoglycerate d
17	88	8.2	436	2 H98058	phosphoglycerate d
18	87	8.1	372	2 G96818	F9K20.3 [imported]
19	87	8.1	399	2 T32923	hypothetical prote
20	87	8.1	1640	2 D86798	prophage p13 prote
21	86.5	8.1	458	2 JN0873	immunophilin p59 -
22	86.5	8.1	544	2 B82048	chaperonin, 60 kd
23	86	8.1	303	2 D83482	probable transcrip
24	86	8.1	339	2 E70710	hypothetical prote
25	85.5	8.0	903	1 C64444	cell division cont
26	85.5	8.0	1396	2 G71529	DNA-directed RNA p
27	85.5	8.0	1396	2 F81666	DNA-directed RNA p
28	85.5	8.0	1441	2 T00335	hypothetical prote
29	85	8.0	292	2 G75084	hypothetical prote

30	85	8.0	878	2	P86795	cation-transportin
31	84.5	7.9	1896	1	RNFF2L	DNA-directed RNA p
32	84	7.9	314	2	A72335	conserved hypothet
33	84	7.9	392	2	G89774	phosphopentomutase
34	84	7.9	513	2	G47788	d-xylose transport
35	84	7.9	600	2	T17436	ATP-binding protei
36	84	7.9	600	2	AF0233	inner membrane ABC
37	84	7.9	775	2	G86696	cation-transportin
38	83.5	7.8	474	2	G97033	beta-glucosidase f
39	83.5	7.8	548	2	AC0043	60 kDa chaperonin
40	83.5	7.8	550	2	S2901	heat shock protein
41	83	7.8	244	2	D64358	ribosomal protein
42	83	7.8	465	2	F97034	aspartyl aminopept
43	83	7.8	633	2	I39585	dnak-type molecula
44	83	7.8	633	2	AE2591	DNAP protein [impo
45	83	7.8	633	2	E97373	dnaj protein (heat
46	83	7.8	1060	2	H86686	DNA-directed DNA p
47	82.5	7.7	279	2	G84435	cobalt transport A
48	82.5	7.7	550	2	S26423	heat shock protein
49	82.5	7.7	566	2	T47934	hypothetical prote
50	82.5	7.7	2339	2	S41121	acetyl-CoA carboxy
51	82	7.7	331	2	T28374	hypothetical prote
52	82	7.7	357	2	H89819	conserved hypothet
53	82	7.7	461	2	G86594	lipoamide dehydrog
54	82	7.7	461	2	C72031	2-oxo acid dehydro
55	82	7.7	513	2	B91185	hypothetical prote
56	82	7.7	513	2	H86031	hypothetical prote
57	82	7.7	636	2	A84252	Htr15 transducer f
58	82	7.7	2104	2	T38774	myosin-3 heavy cha
59	81.5	7.6	547	2	B49203	heat shock protein
60	81.5	7.6	745	1	S75354	probable copper-tr
61	81	7.6	154	2	H82810	bacterioferritin X
62	81	7.6	244	1	R3MX4	ribosomal protein
63	81	7.6	298	2	F69119	cation efflux syst
64	81	7.6	445	2	E70360	glucose inhibited
65	81	7.6	798	2	S29815	N-ras upstream pro
66	81	7.6	1201	2	P86386	hypothetical prote
67	80.5	7.5	311	2	B70451	prephenate dehydro
68	80.5	7.5	374	2	C84040	hypothetical prote
69	80.5	7.5	705	2	T47949	hypothetical prote
70	80.5	7.5	1365	2	T45031	hypothetical prote
71	80	7.5	160	2	T27963	hypothetical prote
72	80	7.5	292	2	E72223	conserved hypothet
73	80	7.5	455	2	F72326	hemolysin-related
74	80	7.5	798	2	S11210	probable unr prote
75	80	7.5	1213	2	T19835	hypothetical prote
76	80	7.5	1487	2	G96827	protein F20Bd7.10
77	79.5	7.4	246	2	F70062	hypothetical prote
78	79.5	7.4	458	1	A42386	hsp 90-binding pro
79	79.5	7.4	488	2	C86183	hypothetical prote
80	79.5	7.4	548	2	AS1045	GroEL protein [imp
81	79.5	7.4	664	2	G98894	protein kinase [im
82	79.5	7.4	698	2	D90541	atp synthase beta
83	79.5	7.4	819	1	C71527	endopeptidase La (
84	79.5	7.4	906	2	A43000	alpha N-catenin -
85	79.5	7.4	1006	2	C70445	ATPase subunit of
86	79	7.4	220	2	A57418	pyruvate kinase -
87	79	7.4	333	2	H82245	glyceraldehyde 3-p
88	79	7.4	338	2	G90203	ethylene-inducible
89	79	7.4	401	2	A33312	ATP synthase beta
90	79	7.4	412	2	AC2152	dolichyl-phosphate
91	79	7.4	486	2	F75045	2-isopropylmalate
92	79	7.4	503	2	C84595	similar to pEARLI
93	78.5	7.4	326	2	T07825	fibrillin homolog
94	78.5	7.4	351	2	I64030	hypothetical prote
95	78.5	7.4	505	2	D89064	isopropylmalate sy
96	78.5	7.4	528	2	S70295	probable membrane
97	78.5	7.4	573	2	AB2401	hypothetical prote
98	78.5	7.4	735	2	T13646	hypothetical prote
99	78.5	7.4	818	2	F89819	endopeptidase [imp
100	78.5	7.4	819	2	E81681	proteinase, Lon fa

ALIGNMENTS

RESULT 1

T49962
Hypothetical protein F8M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49962
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493
A:Accession: T49962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <BEV>
A:Cross-references: EMBL:AL353993; GSPDB:GN00063; ATSP:F8M21.160
A:Experimental source: cultivar Columbia; BAC clone F8M21
C:Genetics:
A:Gene: ATSP:F8M21.160
A:Map position: 5
A:Introns: 200/3; 337/3; 544/3

Query Match 10.2%; Score 109; DB 2; Length 568;
Best Local Similarity 26.7%; Pred. No. 0.73; Mismatches 36; Indels 26; Gaps 8;
Matches 46; Conservative 36; Mismatches 64; Indels 26; Gaps 8;
QY 25 ITEEINKAIDDAIAIEQSEITDPMK--VPDADKFERHVGIVDFKGLAMRN-----IE 77
DB 331 ITISAREVFEDA----YSPTEAVMRLPQKCDKVERDSGLVSFTTLLVPSSRICIL 395
QY 78 ARG---LKQMKQGDANVK--GEGIVKA-----HLIGVHDDIVSMYDIAYKGLDLHPT 128
DB 386 GKGGAIITEMRTKANIIRILGKLNPKVASDDDMQVNFVLLKFSQFLSLRLYF 445
QY 129 THVISDIOQFVVVLS--LEISDEGNITWTSFEVQFANVNVNHIGLSILDPI 178
DB 446 TWSVNSNSNFVQISGLDVAKEALIQITS---RLRNVFDEGAVSALMPV 494

RESULT 2

E90591
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: E90591
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <KUR>
A:Cross-references: GB:AL445566; PID:q14090052; PIDN:CAC13810.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV 6370
A:Genetic code: SGC3

Query Match 9.1%; Score 97.5; DB 2; Length 776;
Best Local Similarity 20.1%; Pred. No. 8.3;
Matches 52; Conservative 37; Mismatches 73; Indels 97; Gaps 10;
QY 22 YDKITEINKAIDDAIA--AIEQSEITDPMKVPD--HADKFE-----RHVGIVD--- 66
DB 437 YNQFQDVQNKILEQAFPKNSWEQTFNPSPDESNTYQDKLDNIGDEDKAKFTGIDELF 496
QY 67 -----FKGELAMRN-----IEARGLKQMKRQGDANVKGEEGVK-----A 101
DB 497 NELKSLEFKKAIARNAIAAIIIGLTLAELKKGESQASRDIVKADFEKEMQHNDNA 556
QY 102 HL-----LIGVHDDIVSMYDIAYKGLDLHPTTHVISDIOQFVVVLSLEIS----- 147

DB 557 HLRKIEKYIFGVNGDEQSKDLDLVKQLSQKDSFTSAVESIH---AKALEFSRKVKAKNV 612
QY 148 DEGNI-----TWTSFEVQFANVNVH 168
DB 613 DENNIKKAVELYLANVWQAIWNKFNKYLEIWTWTSFKKKDQGSDDSMTEIQVINEYVK 672
QY 169 IGLSILDPFVGLSDVLT 187
DB 673 YRSSNVADKLIISVITSTNT 691

RESULT 3

F84241
hypothetical protein Vng0849c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84241
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lia
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: F84241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-449 <STO>
A:Cross-references: GB:AE004437; NID:gi0580415; PIDN:AAG19298.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0849C

Query Match 8.8%; Score 93.5; DB 2; Length 449;
Best Local Similarity 21.1%; Pred. No. 8.7;
Matches 42; Conservative 35; Mismatches 87; Indels 35; Gaps 6;
QY 13 VAVSADPHYDKITEINKAIDDAIAIEQSEITDPMKVPDADKFERHV----- 62
DB 263 VAALADAIRNDVADEDLATIQDAMGGVPFESTRVELTHLQSRVSEIEAYTDALSSFFIDN 322
QY 63 -----GIVDFKGLAMRNIEARGLKQ--MKRQGDANVKGEEGIVKAHLIGVHD 109
DB 323 GTAQDVNLDEDDVADLELSTELSTVDDVAGVEQSVDAADRAALHERVDVDEATVGAVD 382
QY 110 DIVSMEYDIAYKGLDLHPTTHVISDIOQFVVVLSLEISDEGNITWTSFEVQFANVNVH 169
DB 383 D---LESSLDRLRGVDALDERLADTDDVLAALA--DVS--GDVDALEADVAELSEHVD 436
QY 170 GGLSILDPFVGLSDVLT 188
DB 437 -----DDWRDQLSDVFS 449

RESULT 4

G75154
tungsten-containing aldehyde ferredoxin oxidoreductase (aor-1) PAB2085 - Pyrococcus abyssi
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75154
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A:Reference number: A75001
A:Accession: G75154
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-628 <XAM>
A:Cross-references: GB:AJ248284; GB:AL096936; NID:G5457730; PIDN:CAB49318.1; PID:G545782
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2085
C:Superfamily: probable aldehyde ferredoxin oxidoreductase aor-4

[illegible]

60kDa chaperonin XP0615 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R:Accession: F82783
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: F82783
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-547 <SIM>
A:Cross-references: GB:AE003907; GB:AE003849; NID:g9105484; PIDN:AAF83425.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sancelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XP0615
C:Superfamily: chaperonin groEL

Query Match 8.5%; Score 90.5; DB 2; Length 547;
Best Local Similarity 22.8%; Pred. No. 19;
Matches 54; Conservative 43; Mismatches 87; Indels 53; Gaps 12;
QY 13 VAYSADPIHDKTSEINKAIDDA-TAAEQSETIDPMKVPDHADKFERHVGIVDFKGL 71
DB 107 VAAGMNP-----DKRGIDKAVIAAATLTKISK---PSDDKAIAQVATISANSDE 156
QY 72 AMNIEARGLKQMKROGDANVKGEIGVKAHLIGVHDDIVSMYEDLAY----- 120
DB 157 SIGNIIAEMKKVKGEGVITI--EEGTLENEL----DVVEGQFDRGYSPYFINNQQS 210
QY 121 ---KLGLDHPPTH--VISDQDFVVALSLEISDEGNITMSTFVRQFA---NVVNHIGGL 172
DB 211 QIVELDNPTLLHDKKISSVRDLTLVDAVAKESKPLIIAEEVEGEALATLVVNNIRGI 270
QY 173 ----SILDPTFG----VLSDEL-----TAIFOD---TVRKEMTKVLAPAFKRELEK 212
DB 271 IKVCAPKAFGGRKRAMLEDMAVLGTGTVISEVGLSEKATTSHLGKAKKRVSK 327

RESULT 12
T15590
hypothetical protein C24A3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
R:Accession: T15590
R:Favell, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C24A3.
A:Reference number: Z18373
A:Accession: T15590
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-681 <FAV>
A:Cross-references: EMBL:U40424; NID:g1065542; PID:g1065549; PIDN:AAA81461.1; CESP:C24A3
C:Genetics:
A:Gene: CESP:C24A3.1
A:Introns: 35/3; 79/2; 154/3; 193/3; 235/2; 283/3; 321/2; 354/3; 380/2; 422/1; 505/3

Query Match 8.4%; Score 90; DB 2; Length 1558;
Best Local Similarity 23.3%; Pred. No. 75;
Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;
QY 22 YDKITEEINKAIDDAIAIEQSEITIDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGL 81
DB 747 FNTVLDKVEETVETISGESLENNE---MDKAFFSEIFDNVKGIOENLLTGMFRSIEISIV 802
QY 82 QMKRQGDANVKGBEGIVKAHLIGVHDDIVSMYEDLAYKGLDHPPTHVISDQDFVVA 141
DB 803 IQSEKVDLN---ENVVSSIL---DNIEKKEGLLNKLENISSTEGVQETVTEHV-- 851
QY 142 LSLEISDEGNITMSTFV-----RQFANVNVNHIGGLS-----ILDPFGVLSVLTAA--IFQ 191
DB 852 -----EQNV-YVDVDVPAMKQDFGLINLAEAGLKEKFNLEDVFKSESVDITVEEIKD 903
QY 192 DTVRKEMTKVLAPAFKRELEKN 213
DB 904 EPVQKEVEKTVSIIIE-EMEN 924

RESULT 14
A46372
immunophilin FKBP52 - human
N:Alternate names: 56K protein; FK506-binding protein; Hsp56; probable peptidylprolyl cis
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
R:Accession: A46372; A35841; C42576; A45115
R:Peattie, D.A.; Harding, M.W.; Fleming, M.A.; DeCenzo, M.T.; Lippke, J.A.; Livingston, I
Proc. Natl. Acad. Sci. U.S.A. 89, 10974-10978, 1992
A:Title: Expression and characterization of human FKBP52, an immunophilin that associates
A:Reference number: A46372; PMID:93066366; PMID:1279700

A;Accession: A46372.
A;Molecule type: mRNA
A;Residues: 1-459 <PEA>
A;Cross-references: GB:M88279; NID:G186389; PIDN:AAA36111.1; PID:G186390
A;Experimental source: placenta
A;Note: sequence extracted from NCBI backbone (NCBIN:118780, NCBIP:118781)
A;Note: part of this sequence was confirmed by protein sequencing
R;Sanchez, E.R.; Faber, L.E.; Heuzel, W.J.; Pratt, W.B.
Biochemistry 29, 5145-5152, 1990
A;Title: The 56-59-kilodalton protein identified in untransformed steroid receptor complexes.
A;Reference number: A35841; MUID:90335211; PMID:2378870
A;Accession: A35841
A;Molecule type: protein
A;Residues: 'X', 3-14, 'X', 16-21 <SAN>
R;Yem, A.W.; Tomasselli, A.G.; Heinrichson, R.L.; Zurcher-Neely, H.; Ruff, V.A.; Johnson, J. Biol. Chem. 267, 2868-2871, 1992
A;Title: The Hsp56 component of steroid receptor complexes binds to immobilized PK506 and A;Reference number: A42576; MUID:92147620; PMID:1371107
A;Accession: C42576
A;Status: preliminary
A;Molecule type: protein
A;Residues: 'X', 3-18 <YEM>
A;Experimental source: Jurkat cells
A;Note: sequence extracted from NCBI backbone (NCBIP:80691)
R;Wiederrecht, G.; Hung, S.; Chan, H.K.; Marcy, A.; Martin, M.; Calaycay, J.; Boulton, D. J. Biol. Chem. 267, 21753-21760, 1992
A;Title: Characterization of high molecular weight FK-506 binding activities reveals a new A;Reference number: A45115; MUID:93016131; PMID:1363226
A;Accession: A45115
A;Status: preliminary
A;Molecule type: protein
A;Residues: 16-32 <WIE>
A;Experimental source: JURKAT cells
A;Note: sequence extracted from NCBI backbone (NCBIP:116746)
C;Superfamily: human FK506-binding protein FKBP51; BKBP-type peptidylprolyl isomerase h F;50-97/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F;319-352/Domain: tetratricopeptide repeat homology <TR>
F;353-386/Domain: tetratricopeptide repeat homology <TT2>
Query Match 8.3%; Score 88.5; DB 2; Length 459;
Best Local Similarity 24.6%; Pred. No. 22;
Matches 30; Conservative 25; Mismatches 50; Indels 17; Gaps 3;
QY 18 DPVHYDKTEINKAIDDAIAAIQSETI-----DPMKVPDHAD-KFERHV 62
DB 75 DKFSFDLKGKGVKAWDIAIATMKVGEVCHITCKPEYAGSAGSPKIPFNATLVPE--V 132
QY 63 GIVDFKGLAMRNTAARGLKQKQGDANVKGEGIVKAHLIGVHDDIVSMYVDLAYKL 122
DB 133 ELFFPKGEDLTEEDGGIIRIQTRGEGYAKPNEGAIVEALGEGYKDKLFDQRELRFPI 192
QY 123 GD 124
DB 193 GE 194
RESULT 15
T05244
hypothetical protein F18A5.140 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C;Accession: T05244
R;Bevan, M.; Weber, N.; Grueninger, D.; Schmidheini, T.; Bancroft, I.; Mewes, H.W.; Maye submitted to the Protein Sequence Database, February 1999
A;Reference number: 215405
A;Accession: T05244
A;Molecule type: DNA
A;Residues: 1-2137 <BEV>
A;Cross-references: EMBL:AL035528
A;Experimental source: cultivar Columbia; BAC clone F18A5
C;Genetics:

A;Map position: 4
A;Introns: 61/1; 621/3; 691/3; 1438/3; 1476/2; 1608/3; 1713/3; 1799/3
A;Note: F18A5.140
C;Superfamily: Arabidopsis thaliana hypothetical protein F18A5.140
Query Match 8.3%; Score 88.5; DB 2; Length 2137;
Best Local Similarity 21.8%; Pred. No. 1.4e+02;
Matches 45; Conservative 34; Mismatches 82; Indels 45; Gaps 10;
QY 35 DAIAAIEQSETIDPMKVPDHADK----FERHVGIVDFKGLAM--RNIEAR-GLKQMKRQ 87
DB 614 DSHFILELVQADNKNKPEHVEPTLTFILQKTGIVLVNNECGFMPENIRALCDVQGSTKK 673
QY 88 GDANVKGEGI-VKAHLIG--VHDDIVSMEVDLAY-KLGDHLPTT---HVISDIQDF 138
DB 674 GSGGYIGKKGIGKFSVFRVSDAPEIHSNGFHFKFDISGQIGYILPTVVPHPDIESLSM 733
QY 139 VVALSLEISDEG---NIWTMTSFEVROFANVNHGGLSILDPF----- 179
DB 734 LSGRALHLKADGWNCTITLFPRAIDSERTVNH-----EPMFSDLHPSLLFLHRLQC 787
QY 180 ----GVLSDVLTAFQDTRKEMTKV 201
DB 788 IYVRNVLDLSLLVMRKEVVSNNIVKV 813
RESULT 16
B95199
phosphoglycerate dehydrogenase-related protein [imported] - Streptococcus pneumoniae (str C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: B95199
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide n, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I nson, T.; Hickey, B.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: B95199
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-436 <KUR>
A;Cross-references: GB:A9005672; PIDN:AAK75787.1; PID:G14973204; GSPDB:GN00164; TIGR:SP46 A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SPI709
C;Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongation
Query Match 8.2%; Score 88; DB 2; Length 436;
Best Local Similarity 21.8%; Pred. No. 22;
Matches 46; Conservative 34; Mismatches 89; Indels 42; Gaps 8;
QY 30 NKAIDDAIAAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRNTAARGLKQMKRQGD 89
DB 21 NRIAGERISIVEDGEGVTRDRIYATGEWLNRSFSDMTGG---IDVDAPMEQIKHQA 77
QY 90 AN-----VKGEIGVK-----AHLIGVHDDIV-----SMEYDL---AYK 121
DB 78 TAMEEADVIVFVSGKEGIDADEYVARKLYKTKPVILAVNKVDNEMRNDIYDFVALG 137
QY 122 LGDLHPTTHV----ISDIQDFVVALSLEISDEGNITMTSFEVROFANV-----VNHIGGL 172
DB 138 LGEPLFISVHGIGTGDVLDIAIVENLPNEYEEENPDVFKFSLIGRPNVGVKSSLLINAILGE 197
QY 173 S--ILDPFGLVSLDVLTAIFQDTRKEMTKV 201
DB 198 DRVIASPVAGTTRDAIDTHFTDQDQEFMTI 228
RESULT 17
H98065
phosphoglycerate dehydrogenase [imported] - Streptococcus pneumoniae (strain R6)

Matches 47; Conservative 45; Mismatches 73; Indels 62 Gaps 12;

Qy 16 SAQPIHYDKITEINKAIDDAIAAIBOSEITIDPMKVPDHDADKFERHVGIVDFKGLAMRN 75
Db 1142 SMSKAQYDEIVKNAQKORDTISAARKQQT---EVTDKAQK--THDKTVELANSKADKN 1195
Qy 76 IEA-----RGLQMKRQGDANVKYEGIVK-----AHLII----- 105
Db 1196 VKAAAEKQGTVOYTKGFKDSDNLNSIFDINGVNLFLHKGWGNIGHVSLKGLPFGATGTR 1255
Qy 106 GVHDDIVSMYEDLAYKLGDLHPHTHVIS-----DIQDFVVALSL---EISDEGNITMRS 156
Db 1256 GLAQDEALVGESEGFELAH-HPSRGIFAVGQQGPEIRNLKAGTISILPHSNKSGE-FLSLT- 1312
Qy 157 FEVRQFANVNVHGGISILDPIFGVLSVLTALFQDITVRKEMTKVLA 203
Db 1313 -----ANLPAHADGVS-----GFLSDALGWV--KSTYKDVTSVIS 1345

RESULT 21

JN0873
immunophilin p59 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 16-Jul-1999
C/Accession: JN0873; S38475
R/Schmitt, J.; Pohl, J.; Stunnenberg, H.G.
Gene 132, 267-271, 1993
A/Title: Cloning and expression of a mouse cDNA encoding p59, an immunophilin that associates with the catalytic subunit of protein phosphatase 2B
A/Reference number: JN0873; MUID:94040772; PMID:7693350
A/Accession: JN0873
A/Molecule type: mRNA
A/Residues: 1-458 <SCH>
A/Cross-references: EMBL:X70887; NID:g410498; PIDN:CAA50231.1; PID:g410499
A/Experimental source: NIH-3T3 cell
C/Comment: This protein is a part of the cytoplasmic steroid receptor complex and interacts with the catalytic subunit of protein phosphatase 2B
C/Superfamily: human FK506-binding protein FKBP51; BKBP-type peptidylprolyl isomerase hcr
C/Keywords: receptor
F/50-97/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F/319-352/Domain: tetratricopeptide repeat homology <TP1>
F/353-386/Domain: tetratricopeptide repeat homology <TT2>

Query Match 8.1%; Score 86.5; DB 2; Length 458;
Best Local Similarity 23.0%; Pred. No. 31;
Matches 28; Conservative 27; Mismatches 50; Indels 17; Gaps 3;

Qy 18 DPHYDKITEINKAIDDAIAAIEQSETI-----DPMKVPDHDAD-KFERHV 62
Db 75 DKPSFDLKGGEVIAKWDIAVATMKVGEVCHITCKPEYAVGAAGSPFKIPNATLYPE--V 132
Qy 63 GIVDFKGLAMRNIEARGLKQMKRQGDANVKYEGIVKHAHLIGHDDIVSMYEDLAYKL 122
Db 133 ELPEFKGEDITEBEDGGITRRIRTRGEGYARPNDGAMVEALEGYHKDLFDQRELCFEV 192
Qy 123 GD 124
Db 193 GE 194

RESULT 22

B82048
chaperonin, 60 kD chain VC2664 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: B82048
R/Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.J.;
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 405, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: B82035; MUID:20406833; PMID:10952301
A/Accession: B82048
A/Status: preliminary
A/Molecule type: DNA

Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70710
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <COL>
A:Cross-references: GB:279701; GB:AL123456; NID:93261635; PIDN:CAB02032.1; PID:g1524228
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1482c
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1073

Query Match 8.1%; Score 86; DB 2; Length 339;
Best Local Similarity 28.3%; Pred. No. 24;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;

QY 33 IDDAIAIEQSETIDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGLKQMKR----- 86
Db 187 LDPAAVADALIQTADKADVEPLIERGRGKMAALDLVGGASPKETWRLLL 246

QY 87 -----QGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKLGLDHPHT--HVISDI 135
Db 247 LIRAGFRPQTQIAVRNMGWAEHLDMGQDIKVAEYD-----GDHLLTSRYHYRKDI 301

RESULT 25
C64444
cell division control protein CDC48 homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: C64444
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: C64444
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-903 <BUL>
A:Cross-references: GB:U67557; GB:L77117; NID:g1591777; PIDN:AAB99153.1; PID:g1591785; T
C:Genetics:
A:Map position: REV1097124-1094413
C:Superfamily: transitional endoplasmic reticulum ATPase; FtsH/SEC18/CDC48-type ATP-bind
C:Keywords: ATP; nucleotide binding; P-loop
F;193-400/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VAT1>
F;220-227/Region: nucleotide-binding motif A (P-loop)
F;466-674/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VAT2>
F;493-500/Region: nucleotide-binding motif A (P-loop)

Query Match 8.0%; Score 85.5; DB 1; Length 903;
Best Local Similarity 25.0%; Pred. No. 85;
Matches 53; Conservative 30; Mismatches 90; Indels 39; Gaps 11;

QY 23 DKITEENKAIID-----ATAAIEQSSTIDPMKV-PDHADKFERHVGIVDFKGLAM 73
Db 571 DKVNVQLLTDLGNEEPKQDVVIAATNRPIIDPALLRFGRLDRVIL-VPVPEKARLDI 629

QY 74 RNIEARGLKQMKRGDANV-----KGEIGIVKAHL-----LIGVHDDIVSMEYDLAYK 121
Db 630 FKIHTR---SNMLAEDVNLBELAKKTGTYTGADIEALCREAAMLAVERSI-GKPDVIEVK 685

QY 122 LGDLHPTTHVISDIDQFVWLSLEISDEGNIT--MTSFEVRQFANVNVHIGLSILDPIF 179
Db 686 LREL---INVLOSISGTFRAAAVELNSVIKATKERSAEAGEFSLEKNAIG-----KII 736

QY 180 GVLSDLVLTALFQDTPVRKEMTKVLAPAKRELE 211
Db 737 SVLSPAKEKI--EAVEKEIDKFLEVINKEELK 766

RESULT 26

F81529

DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain - Chlamydia trachomatis (serotype D, C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 18-Jun-1999
C:Accession: G71529; C36135
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trach A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: G71529
A:Molecule type: DNA
A:Residues: 1-1396 <ARN>
A:Cross-references: GB:AE001304; GB:AE001273; NID:G3328730; PIDN:AAC67907.1; PID:G3328731
R:Engel, J.N.; Pollack, J.; Malik, F.; Ganem, D.
J. Bacteriol. 172, 5732-5741, 1990
A:Title: Cloning and characterization of RNA polymerase core subunits of Chlamydia trach A:Reference number: A36135; MUID:91008945; PMID:2211507
A:Accession: C36135
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra A:Molecule type: DNA
A:Residues: 124-128, 'EK', 131-205, 'D', 207-239 <ENG>
C:Genetics:
A:Gene: rpoC
C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain
C:Keywords: nucleotidyltransferase; transcription

Query Match 8.0%; Score 85.5; DB 2; Length 1396;
Best Local Similarity 24.0%; Pred. No. 1.5e+02;
Matches 50; Conservative 33; Mismatches 66; Indels 59; Gaps 12;

QY 43 SETIDPMKVPDHADKFERHVGIVDFKGLAMRNIE-----EARGLKQ-----MKRQ 87
Db 1145 ALBVEARKPEDAAD-IAKIDGVDFKGIQKXKRIILVDEITGMEEHLISLTKHLIVQR 1203

QY 88 GDANVKGK--EGIVKAHLII--GVHD--DIVSMEYDLAYKLGLDHPHTTHVISDIDQFV 139
Db 1204 GUSVINGQQLDGLVVPHEILICGVRLEQKVLNVEQVETRLQGV-----DINDKH 1255

QY 140 VAL-----SLEISDEGNITM---TSFEVRQFANV---VNHIGG-----LSILD 177
Db 1256 VBIIVRQMLQKVRITDPGDTLLFGEDVDKKEFYENRRTEEDGKPAQAVPVLGITKA 1315

QY 178 IGVLSDLVLTALFQDTPVRKEMTKVLAPA 205
Db 1316 SLGTESFISAASFQDT-----TRVLTD 1338

RESULT 27

F81686

DNA-directed RNA polymerase, beta' chain TC0588 [imported] - Chlamydia muridarum (strain C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: F81686
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, I. C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: F81686
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1396 <TET>
A:Cross-references: GB:AS002328; GB:AS002160; NID:g7190627; PIDN:AAF39420.1; PID:g719062 A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0588
C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain

Query Match 8.0%; Score 85.5; DB 2; Length 1396;

Best Local Similarity 23.6%; Pred. No. 1.5e+02;
Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12;
QY 43 SETIDPMKVPDHAKFERHVGIVDFPKGLAMRNI-----EARGLKQ-----MKRQ 87
Db 1145 AELVEARKPEDAAD-IAKIDGVDFKGIQKQKRILVVRDEVGTGMBEHLISLTGHLIVQR 1203
QY 88 GDANVKGB---EGIVKAHLII---GVHD--DIVSMEDVLAAYKLGDLHPTTHVISDIDQFV 139
Db 1204 GDSVIKQQLDGLVVPREIIEICGVELQKLYLNEVQEVYRLQGV-----DINDKH 1255
QY 140 VAL-----SLEISDEGNITM---TSPEVRQFANV---VNHIGG-----LSILDP 177
Db 1256 IEIIVRQMLQKVRITPDGDTILLFGEDVDKKEFEYENRTEEDGKPAQAVPVLGITKA 1315
QY 178 IFGVLSVLTALFQDTRVKEMTKVLAPA 205
Db 1316 SLGTESFISAASFQD-----TRVLTD 1338
RESULT 28
T00335
Hypothetical protein KIAA0564 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00335
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00335
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1441 <NAG>
A:Cross-references: EMBL:AB011136; NID:g3043651; PIDN:BA025490.1; PID:g3043652
A:Experimental source: brain; clone HH1811
C:Genetics:
A>Note: KIAA0564

Query Match 8.0%; Score 85.5; DB 2; Length 1441;
Best Local Similarity 22.0%; Pred. No. 1.5e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;
QY 20 IHYDKI--TEEINKAIDAAIEQSETIDPMK-----VDPHADKFERHV-----GIVDF 67
Db 343 LHRDTTGTTLQPSVKDGLIVEDSPLVKAVKLGHLIVVDEADKAPTNTVTCILKTLVE- 401
QY 68 KGLAM--RNTEARGLKQKQGDANVKREGIVKAH-----LLIGVHDDIVSMEDLA 119
Db 402 NGEMLADGRIVA-----NSANNVGNVNVVTHPDFTMVLNRPQFPFLGNDFF 452
QY 120 YKLGDLHPTTHVISD-----IQDFVVALS--LEISDEGNITMVS 156
Db 453 GTLGDIF-SCHAVDNPKPSELEMLRQYGNVPPEPIQLKLVAAFGELSLADQGIINY-P 510
QY 157 FEVRQFANVNHIGGLSILDPFGVLSVLTALFQ-DIVRKEMTKVL 202
Db 511 YSTREVNIVRHLQKF-----PTEG-LSSVVRNVDFDSYNNDMREIL 552

RESULT 29
G75084
Hypothetical protein PAB0711 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75084
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: G75084
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-232 <KAW>
A:Cross-references: GB:AJ249286; GB:AL096836; NID:g5458366; PIDN:CAB49980.1; PID:g545849;
A:Experimental source: strain Orsay
C:Genetics:
C:Gene: PAB0711
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1169
Query Match 8.0%; Score 85; DB 2; Length 292;
Best Local Similarity 20.7%; Pred. No. 23;
Matches 44; Conservative 31; Mismatches 62; Indels 76; Gaps 7;
QY 29 INKAIDD-----AIAAIEQSTIDPMKVPDHADKFERHVGIVDFK---- 68
Db 11 INKVIENPRVIFPLVIVLIFSVPFLAFLEKSNIKPLNFEEAGVIEKHGAISDMKPLNL 70
QY 69 -----GELAMNIEA-----RGLQKQKQGDANVKREGIVKAHLIGVHDDIVSNEY 116
Db 71 KLLLLVGLLQLLLSAVQYIIHYVKTGTGTGWEAFKGLNENVIOMFL----- 118
QY 117 DLAYKLGDLHPTTHVISDIDQFVWALSLEISDEGNITMTSFEVRQF-ANVNVNHIGL--S 173
Db 119 -----NVISYL-----IVLVAFLIAIFPMPMIIVGIGILES 149
QY 174 ILDPFGVLSVLTALFQDTRVKEMTKVLAPAF 206
Db 150 TGAIVFGILLVILSLGLIIGSFAIGMTSVIVPAY 182
RESULT 30
F86795
cation-transporting ATPase YoaB [imported] - Lactococcus lactis subsp. lactis (strain ILJ
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: F86795
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A:Reference number: AB6625; MUID:21235186; PMID:11337471
A:Accession: F86795
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-878 <STO>
A:Cross-references: GB:AE005176; PID:g12724350; PIDN:AAK05464.1; GSPDB:GN00146
A:Experimental source: strain ILI403
C:Genetics:
A:Gene: yoaB
Query Match 8.0%; Score 85; DB 2; Length 878;
Best Local Similarity 24.7%; Pred. No. 90;
Matches 39; Conservative 33; Mismatches 44; Indels 42; Gaps 8;
QY 72 AMRNIARGLKQKQGDANVKREGIVKA-HLLIGVHDDIVSMEDVLAAYKLGDLHPTTH 130
Db 106 ALRQMSAPSAKVL-----NGKTSIPARELVVG---DIVSLE-----AGDFIPADG 149
QY 131 VTSDIQDFVVALSLEISDEGNITMTSFEVRQFANV-----NHIGGLSILDPFG 180
Db 150 RLIDVQNLV-----EGMLTGESEPEKESDVIEGVALGDRKNVVFSSLV--VYG 200
QY 181 VLSVLTALFQDTRVKEMTKVLAPA-----FKRELEK 212
Db 201 RADFLVTAIEQTEIGKTAQMLETAEAKQTPLQOKLEK 238

RESULT 31
RNFP2L
DNA-directed RNA polymerase (EC 2.7.7.6) II 215K chain [validated] - fruit fly (Drosophila
C:Species: Drosophila melanogaster
C>Date: 04-Dec-1986 #sequence_revision 31-Mar-1993 #text_change 28-Jul-2000
C:Accession: S04457; A00693; B27677; S60151
R:Jokrest, R.S.; Weeks, J.R.; Zehring, W.A.; Greenleaf, A.L.
Mol. Gen. Genet. 215, 266-275, 1989
A:Title: Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosoph

A;Reference number: S04457; MUID:89218930; PMID:2496296

A;Accession: S04457

A;Molecule type: DNA

A;Residues: 1-1896 <JOK>

A;Cross-references: EMBL:M27431; NID:g158331; PIDN:AAA28868.1; PID:g158332

R;Biggs, J.; Seales, L.L.; Greenleaf, A.L.

Cell 42, 611-621, 1985

A;Title: Structure of the eukaryotic transcription apparatus: features of the gene for b

A;Reference number: A00693; MUID:85282618; PMID:2992806

A;Accession: A00693

A;Molecule type: DNA

A;Residues: 1-318, 'GVAKV', 325-449, 'G', 451-454, 'RCTT', 459-462, 'VTGESVASS' <BIG>

A;Cross-references: EMBL:M11798

A;Note: this sequence has been revised in reference S04457

R;Allison, L.A.; Wong, J.K.C.; Fitzpatrick, V.D.; Moyle, M.; Ingles, C.J.

Mol. Cell. Biol. 8, 321-329, 1988

A;Title: The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomy

A;Reference number: A93104; MUID:88094402; PMID:3122024

A;Accession: B27677

A;Molecule type: DNA

A;Residues: 1441-1484, 'I', 1527-1889 <ALL>

A;Cross-references: EMBL:M19537; NID:g158147; PIDN:AAA28827.1; PID:g158148

R;Peterson, G.; Song, D.; Ruehle-Doerr, B.; Oldenburg, I.; Bautz, E.K.F.

Mol. Gen. Genet. 249, 425-431, 1995

A;Title: Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragme

A;Reference number: S60151; MUID:96133682; PMID:8552047

A;Accession: S60151

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 778-827 <PET>

C;Genetics:

A;Gene: RPO21; R01215

A;Cross-references: FlyBase:FBgn0003277

A;Map position: X,10C, X 35.7

A;Introns: 27/3; 775/3; 1526/1

C;Function:

A;Description: EC 2.7.7.6 [validated, MUID:88094402]; essential for proper initiation of

C;Superfamily: human DNA-directed RNA polymerase II largest chain

C;Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc fing

F;67-83/Region: zinc finger CCH motif

F;1581-1883/Region: 7-residue repeats

F;349/Binding site: ATP/GTP (Lys) #status predicted

Query Match 7.9%; Score 84.5; DB 1; Length 1896;

Best Local Similarity 22.7%; Pred. No. 2.5e+02;

Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;

Qy 17 ADPHYDKITEINKAIDDAIAAEQSEITDPMKVPDHA--DKFERHVGIV-----DFKG 69

Db 686 ADPQTYNEIOQAIAKKAADDVINVIQAKHMELEPTGNTLRQTFENKVNRLNDAHDKTG 745

Qy 70 ELAMRNI--EARGLKQMKRQ-----GDANVKGBE--GIVKAHLILIGVHD 109

Db 746 GSAKSLTEYNLAKMNVVSGSGKNINISQVIACVGGQNVGKRIIPYGFKRKRLPHFIKD 805

Qy 110 D-----IVSMVYDLAYKGLDHPHTT---HVISDIQDFVVALSLEISDEGNI-----T 153

Db 806 DYGPESRGFVENSY-----LAGLTFSEFYHAMGG-REGLIDTAVKTAETGYIQRLIKA 859

Qy 154 MTSFEVRQFANVNVHIGLSLTL---DPIFGVLSD----- 184

Db 860 MESVNVYDGVNRVNSVGQLIQRLIGEDGLCGELVEFQNMFTVKNLSFKPKRPFDSNE 919

Qy 185 -VLTAIFQDVTYRKEMT 199

Db 920 RLKKKVFDDVIKEMT 935

RESULT 32

A72335

conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C;Accession: A72335

R;Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinp, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, R.; Nelson, K.E.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: A72335

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-314 <ARN>

A;Cross-references: GB:AE001747; GB:AE000512; NID:g4981304; PIDN:AAD35882.1; PID:g498132;

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM0800

Query Match 7.9%; Score 84; DB 2; Length 314;

Best Local Similarity 23.5%; Pred. No. 31;

Matches 48; Conservative 32; Mismatches 76; Indels 48; Gaps 10;

Qy 23 DKITEINKAIDDAIAA--IEQSEITDPMKVPDHDADKPERHV-----GIVDFKGELA 72

Db 122 DSLAMVFRAGADAVIAEGMESGGHIGVTTFVLVNVKVSINVPIVIAAGGIADGRGMAA 181

Qy 73 MENIEARGLKQMKR-----QGDANVKGBEGIVKXAHLLIGVHDDIVSMYDLYAYKGLDHP 127

Db 182 APALGAEAVQMGTRFVASVESDVHPVYKEKIVKA---SIRDVTVT-----GAKLGG--HP 230

Qy 128 T-----THVISDIQDFVVALSLEISD-----EGNITMTSFEVRQFANVNVHIGGL 172

Db 231 ARVLRTPFARKIQEMEFENPMQAEMLVGSURRAVVEGDLRGSPFMVQCSAGLIDEI--- 287

Qy 173 SILDFIGVLSDLTAIFQDVTYRK 196

Db 288 ---KPVKQIIEDILKE--FKETVEK 307

RESULT 33

G89774

phosphotomutase [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C;Accession: G89774

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: G89774

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-392 <KUR>

A;Cross-references: GB:BA000018; PID:g13700055; PIDN:BA041354.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: drm

C;Superfamily: phosphotomutase

Query Match 7.9%; Score 84; DB 2; Length 392;

Best Local Similarity 24.1%; Pred. No. 40;

Matches 45; Conservative 35; Mismatches 85; Indels 22; Gaps 10;

Qy 14 AVSADPIHYDKITE-EINKAIDDAIAAEQSEITDPMKVPDHDADKFERHVGIVDFKGELA 72

Db 65 AVEQPEAYTKLSEASVGKDTMTGHWELNIMQFFKYPNGFPFELIQIIEEMTGKRV 124

Qy 73 MENIEARGLKQMKRQGDANVKGBEGIVKXAH-----LLIGVHDDIVSME--YDLYAYKGLD- 125

Db 125 VANKPASGTQIDWEHGHOMKTDGLIVYTSADPVLQIAAHEDIIPLEELYDICEKVEILT 184

Qy 126 HPHTHVISDIQDFVVALSLEISDEGNITWTS-----FEVRQFA-NVNVNHI--GGLSILQPI 178

Db 355 EGDGHALHVSITPPAASMSALVGASGAGTITVTKLMR-YADPQQQISIGGVDIRRLT 413
QY 179 FGVLSDLVTAIFOD 192
Db 414 PEQLNSLISVVFOD 427

RESULT 37
G86696
cation-transporting ATPase yfgQ [imported] - Lactococcus lactis subsp. lactis (strain IL11403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86696
R:Bohlin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissbach, J.; Ehrlich, S.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis subsp. lactis IL11403
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-775 <STO>
A:Cross-references: GB:AB005176; PID:g12723465; PIDN:AAK04673.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yfgQ

Query Match 7.8%; Score 84; DB 2; Length 775;
Best Local Similarity 19.8%; Pred. No. 93;
Matches 50; Conservative 37; Mismatches 103; Indels 62; Gaps 6;

QY 22 YDKITEIKKAIDAAIAAISOSETIDPMKVPDHDKFERHVGIVDPKGL-AMRN 75
Db 397 YEKVSSINHLLEGFRVIVLAGTKE--KIYDQLNGAYALGVVLANPIRENAKSTFNY 454
QY 76 IEARGLKMKRQGD-----ANVKGEGIVKAHL----- 104
Db 455 FAEQGVNIKVISGDPQTVSAVAKRAGITGAERFIDANLLKTKEDLDQAVESYTVFGRVT 514
QY 105 -----IGVHDDIVSMEYD-----LAKGLDLHPTHTVTSIDQDFVVALSLISDE 149
Db 515 PDQKRRLVQALKRKDHVTAMTGQVNDILAKWSADCSIAMSGSDAATQVAVQLDSDP 574
QY 150 GNITMTSFEVRQFANVNHIGLSILDPFGLSDVLTAIFQDVRKEMTK-----V 201
Db 575 GHTQVTVTEGRVNVNQRSALLFLVKNLSIILAIISAIFVFTYPLQASQLSLISLFTI 634
QY 202 LAPAKRELEKN 213
Db 635 GIPGFLLSLEEN 646

RESULT 38
G97033
beta-glucosidase family protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: G97033
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <KUR>
A:Cross-references: GB:AB001437; PIDN:AAK79058.1; PID:g15024000; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1084
C:Superfamily: Agrobacterium beta-glucosidase

Query Match 7.8%; Score 83.5; DB 2; Length 474;
Best Local Similarity 23.3%; Pred. No. 55;
Matches 34; Conservative 22; Mismatches 65; Indels 25; Gaps 3;

QY 53 DHADKFERHVGIVDFKGLAMRNTEARGLKQMKQGDANVKEE---GIVKAHLIGVHD 109
Db 55 DHYHFKEDVALMAEMGLKAYRFSWSRITVPKRGKYNLKGLEFYKVLVDLLKYNIEP 114
QY 110 DIVSMEYDLAVKLGDLH---PTHTVISIDQDFVVALSLISDEGNITMTSFEVROFAN-- 164
Db 115 VLTLYHMDLPQALQDLGGFESRNVISDFENVCTILFNFKDKVKYWIITFEQNVFTNLG 174
QY 165 -----VVNHIGLS 173
Db 175 YRCVAVHPFPIKIDIKKYLANHIVNLA 200

RESULT 39
AG0043
60 kDa chaperonin [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Eil, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, F.; Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-548 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89210.1; PID:g15978449; GSPDB:GN00175
C:Genetics:
A:Gene: groEL
C:Superfamily: chaperonin groEL

Query Match 7.8%; Score 83.5; DB 2; Length 548;
Best Local Similarity 18.1%; Pred. No. 66;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPHYDKITEINKAIDDA-IAAISOSETIDPMKVPDHDKFERHVGIVDPKGL 71
Db 107 VAAGNPNM-----DLKRGIDKAVIAAVEE---LKLSPVPCSDSKAIAQVGTISGSDS 156
QY 72 AMRNTEARGLKQMKQGDANVKEGIVKAHLIGVHD-----IVSMEYDLA----- 120
Db 157 TVGELIAQAMEKV-----GKSGVITVBEGSLQDELVDVVEGMQDFRGLSPYFINK 207
QY 121 -----KLGDLHPTHTVISIDQDFVVALSLISDEGNITMTSFEV 159
Db 208 PETGSIELESPPILLADKKISNIREMLFVLEAVAKAGKPLLIADVEGEALATL----- 262

QY 160 ROFANVVNHIGL-----SILDPFGLSDVLTAIFQDVTVRKEMTKVLAPAKRELEK 212
Db 263 -----VNTMREGIVKVAAVKAPFG---DRRKAMLQDIATLTAGTVISEEIGLELEK 311

RESULT 40
S52901
heat shock protein 60K - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
C:Accession: S52901
R:Autenrieth, I.; Noll, A.
A:Description: Characterization of protective T-cell epitopes of Yersinia enterocolitica
A:Reference number: S52901
A:Accession: S52901
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-550 <AUT>

A;Cross-references: EMBL:X82212; NID:G695775; PIDN:CAA57694.1; PID:G695776
C;Superfamily: chaperonin groEL

Query Match 7.8%; Score 83.5; DB 2; Length 550;
Best Local Similarity 18.1%; Pred. No. 66;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;
QY 13 VAVSADPHYKITEINKAIDDA-IAAIEQSETIDPMKVPDHPADKFERHVGIVDFGEL 71
DB 107 VAAAGNPM-----DLRGIDKAVIAAEE---LKLSPVCSKAIQAQVGTISANS 156
QY 72 AMRNTAAGLQKMKRQGDANVKGEIGVKAHLIGVHDD---IVSMEYDLAY----- 120
DB 157 TVGELIAAMEKV-----GREGVITVEGSLQDELVDVVGMDPDRGLSPYFINK 207
QY 121 -----KLGLDHPHTTVISIDIQFVVALSLEISDEGNITMTSFEV 159
DB 208 PETGSIELESFILLADKKISNIREMLPVEAVAKAGKPLIIIAEDVEGEALATL----- 262
QY 160 RQFANVNHIGL-----SILPIFGVLSDLTAIFQDVTVRKEMTKVLAFAFKRELEK 212
DB 263 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEEIGLELEK 311

RESULT 41

D64358
ribosomal protein S4E - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
R;Accession: D64358
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Rison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Science 273, 1058-1073, 1996
A;Authors: Kaibe, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; et al.; Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: D64358
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-244 <BUL>
A;Cross-references: GB:U67497; GB:L77117; NID:G2826284; PIDN:AAB98457.1; PID:G1591170; T
C;Genetics:
A;Map position: FOR414312-415046
C;Superfamily: rat ribosomal protein S4

Query Match 7.8%; Score 83; DB 2; Length 244;
Best Local Similarity 21.5%; Pred. No. 27;
Matches 42; Conservative 30; Mismatches 87; Indels 36; Gaps 6;
QY 6 LIAAVAFVAGSADPHYKITEINKAIDDAIAIEQSETIDPMKVPDHPADKFERHVGIV 65
DB 45 LLLIVRDILKYADNAREAKIKMGKVLVDGRVVRKEELPVGLMDVWSLPDANENYRVLF 104
QY 66 DFKGELAMNTEARGLQKMKRQGDANVKGEIGVKAHLIGVHDD---DIVSMEY 116
DB 105 DRKGRIKUPTEPNPVCKLKNKTVIKG-----GHQLMLHGRNIVKVSPTKAE 158
QY 117 DLAYKLGLDHPHTTVISIDIQFVVALSLEISDEGNITMTSFEVQFANVV--NHIGLSI 174
DB 159 DV-YKTGD-----TLISIQEIKAHIPVEGVGLAYITGGKHVGDFAK 201
QY 175 LDPI--FGVLSDLT 187
DB 202 IVEIERRGIVPDIVT 216

RESULT 42

F97034
aspartyl aminopeptidase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: F97034

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: F97034
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-465 <GUR>
A;Cross-references: GB:AE001437; PIDN:AAK79065.1; PID:G15024007; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1091

Query Match 7.8%; Score 83; DB 2; Length 465;
Best Local Similarity 28.4%; Pred. No. 59;
Matches 42; Conservative 27; Mismatches 49; Indels 30; Gaps 10;
QY 74 RNIE---ARGLQKMKRQGD---ANVK-----GEEGIVKAHLIGVHDDI--VSMEY 116
DB 58 RNIEDILAKG--ETLKEGDKVYANNRGKGLIMFLIGKEPLTYTGFKILGAHIDSRLDLQK 115
QY 117 DLAYKLGLDHP--TTHVISIDIQFV--VALSLEISDEGNITMTSFEVQFANVVNHIGLSI 174
DB 116 NPLYEDTDLAMLETHYGGIKYQWVTLFLAI--HGVI-----VKDGTIVNVCVBDD 167
QY 175 LDPIFGVLSDLTAIFQDVTVRKEMTKVL 202
DB 168 NNPVFGV--SDILVHLASOLEKASKVI 194

RESULT 43

I39585
dnaK-type molecular chaperone dnaK - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 20-Aug-1999
C;Accession: I39585
R;Segal, G.; Ron, E.Z.
J. Bacteriol. 177, 5952-5958, 1995
A;Title: The dnaK operon of Agrobacterium tumefaciens: transcriptional analysis and evolution
A;Reference number: I39585; MUID:96011387; PMID:7592349
A;Accession: I39585
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-633 <RES>
A;Cross-references: EMBL:X87113; NID:G1027503; PIDN:CAA60592.1; PID:G1027504
C;Genetics:
A;Gene: dnaK
C;Function:
A;Description: involved in protein folding and assembling/disassembling of protein complex
C;Superfamily: heat shock protein 70
C;Keywords: ATP; molecular chaperone

Query Match 7.8%; Score 83; DB 2; Length 633;
Best Local Similarity 28.3%; Pred. No. 86;
Matches 28; Conservative 16; Mismatches 27; Indels 28; Gaps 4;
QY 23 DKITEINKAIDDAIA---AIEQSETIDPMKVPDHPADKFERHVGIVDFKGLAMNTEA 78
DB 550 DKVSETDRKAIEDAIAKLTAVERAAE-----PDADDIQAKTQTLMEVSMKLGAIYE 602
QY 79 RGLQKMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYD 117
DB 603 -----QQAEGDASAEK-----DDVVDADYE 624

RESULT 44

AE2591
DNAK Protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AE2591
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.

y	73	MKNLEAKGL---	QMAKQGGANVKGEGVAAAGRLIIGVAD-DI	TSMEIDLRKAUGD-	167
b	162	VLDEFTAGLDPVGASIKMLLYDLNKGK	-----WTIIISTHEDVDLPVPVADKVVYVYDG	215	
y	135	--LHPTT--HVISDIQDF	-----VVALSLEI--SDEGNITM----	TSPVROFANVVN	167
b	216	KILKEGTPKEVFSDEVIRKANURLPVAHLIEILNKKDNIP	EWGFTIGEVR--NIVN	273	
y	168	HI 169			

QY	136	-QDFVVALSLEISDEGNITMTSFVRQPANVVNHIGGL-----SILDPIFGVLSDVL	186
Db	452	QESFDRLPKTKDSE-----VIHVGLKADAMSNTYSLLDVSYGIL---L	492
QY	187	TAIFQDTV 194	
Db	493	QIFFSEVV 500	
RESULT 50			
S41121	acetyl-CoA carboxylase (EC 6.4.1.2) - human		
C;Species:	Homo sapiens (man)		
C;Date:	19-May-1994	#sequence_revision	10-Nov-1995 #text_change 11-Jan-2002
C;Accession:	S41121		
R;Ra,	J.; Daniel, S.; Kong, I.-S.; Park, C.K.; Tae, H.J.; Kim, K.H.		
Eur.	J. Biochem.	219,	297-306, 1994
A;Title:	Cloning of human acetyl-CoA carboxylase cDNA.		
A;Reference number:	S41121;	MUID:	94139704; PMID:7905825
A;Accession:	S41121		
A:Molecule type:	mRNA		
A;Residues:	1-2339 <HAJ>		
A;tross-references:	EMBL:X68968; NID:G452315; PIDN:CAA48770.1; PID:G452316		
C;Superfamily:	human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-		
C;Keywords:	biotin binding; ligase		
F;120-620/Domain:	biotin carboxylase homology <BCH>		
F;747-819/Domain:	lipoyl/biotin-binding homology <LPB>		
F;786/Binding site:	biotin (lys) (covalent) #status predicted		
Query Match	7.7%;	Score	82.5; DB 2; Length 2339;
Best Local Similarity	24.2%;	Pred. No.	4.7e+02;
Matches	39; Conservative	32; Mismatches	49; Indels 41; Gaps 7;
QY	45	TIDPMKVPHADKFRHVGIVDFKGEL-----AMENIEARGLKQMKGQ----	88
Db	2091	TINPLCIEMYADK-ESRGGVLEPEGTVEIKRKEDLIKSMRRIDPAYKKLMQLGEPLDS	2149
QY	89	DANVKGEGIVKAHLIGVHDHDIVSMEYDLAYKLGD LHPT-----THVISDIQDFVVAL	142
Db	2150	KDKRDKEGRKAR-----EDLLLIYHQVAQVFADFHTFGRMLEKGVISDILEWKTAR	2204
QY	143	S-----LEISDRGNITMTSFVRQPANVVNHIGGLSIL	175
Db	2205	TFLYWLRLRLLEDQ-----VKQEILQASGELSHVHIQSML	2240
RESULT 51			
T28374	hypothetical protein 213 - Melanoplus sanguinipes entomopoxvirus		
C;Species:	Melanoplus sanguinipes entomopoxvirus		
C;Date:	21-Jan-2000	#sequence_revision	21-Jan-2000 #text_change 16-Feb-2001
C;Accession:	T28374		
R;Afonso, C.L.;	Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.		
J. Virol.	73,	533-552,	1999
A;Title:	The genome of Melanoplus sanguinipes entomopoxvirus.		
A;Reference number:	Z20484; MUID:99102612; PMID:9847359		
A;Accession:	T28374		
A>Status:	preliminary; translated from GB/EMBL/DBDJ		
A:Molecule type:	DNA		
A;Residues:	1-331 <AFO>		
A;tross-references:	EMBL:AF063866; NID:G4049647; PIDN:AAC97703.1; PID:G4049743		
C;Genetics:			
A;Note:	MSV213		
C;Superfamily:	Melanoplus sanguinipes entomopoxvirus hypothetical protein 213		
Query Match	7.7%;	Score	82; DB 2; Length 331;
Best Local Similarity	22.5%;	Pred No.	47;
Matches	45; Conservative	30; Mismatches	53; Indels 72; Gaps 11;
QY	20	IHYDKITE-EINKATDDIAI-----ATEQSETIDPMKVPHADKFRHHVGIQFKGELA	72
Db	62	ICYDKISLSKNSIADKITITFKDNGIYVEELI-----FENNINIIDPK-----	106

```
QY 73 MRNIEARGLKQMRQGDANVKGEGIVKHAHLIGVHDDIVSMEYDLAYKGLDLPHTHTVI 132
Db 107 --YIMOSIQINTPSNN--NNPASARI--DDI--SDINETDNI 145
QY 133 SDIQDFVVALSLEISDEGNITM--TSFEVRQFANVNNHIGLSILDPIF 179
Db 146 SE-----KLEKXHSGLDITGVTIHDKFTLNTNTYKYRINPNMYNKVGA-----VF 193
QY 180 GVLSDVLTAIFQDTRKEMT 199
Db 194 KIKLDVLT-----EFIKLMT 209

RESULT 52
H89819
C:Species: Staphylococcus aureus
C:Conserved hypothetical protein SA0485 [imported] - Staphylococcus aureus (strain N315)
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: H89819
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: H89819
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <KUR>
A:CROSS-references: GB:BA000018; PID:gl3700417; PIDN:BA041715.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0485
C:Superfamily: conserved hypothetical protein yacL

Query Match 7.7%; Score 82; DB 2; Length 357;
Best Local Similarity 23.9%; Pred. No. 51;
Matches 3; Conservative 25; Mismatches 51; Indels 26; Gaps 6;

QY 89 DANVKGEGIVKHAHLIGVHDDIVSME-----YDLAYKGLDLPHT--THVISDI 135
Db 183 DGNILIPQGVNELQIVADNSDVSKRGKGLDILNELYDLPYPTKVIHPTKTH--SDI 240
QY 136 QDFVVALSLEISDEGNITMSTFEVRQFANVNNHIGLSILDPIFGVLSVLTAFQDTR 195
Db 241 DTMLLKLAKQY--HASIITTFDNL-----NKVCHVHGKALN-----VNDLSEAIPNVHQ 289
QY 196 KEMTKVLAPAFKRE 209
Db 290 GDQLHILLTWGKE 303

RESULT 53
G86594
lipamide dehydrogenase [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 20-Apr-2001
C:Accession: G86594
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: G86594
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <STO>
A:CROSS-references: GB:BA000008; NID:g8979207; PIDN:BAA99041.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: lpdA
C:Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology
C:Keywords: redox-active disulfide
F:42-47/Disulfide bonds: redox-active #status predicted
```

```
Query Match 7.7%; Score 82; DB 2; Length 461;
Best Local Similarity 20.4%; Pred. No. 70;
Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;

QY 23 DKITEE-----INKALDDAIAIEQSETIDPMKVPDHDADKFE-----RH 61
Db 219 NKFTKQGRILTKA---SISAIESQNVRTVNDQVEEFYDLVAIGRQFNNTASIGLDN 275
QY 62 VGIV--DFKGELAMRNIEARGLKQMKRQD-----ANVKGEGIVKHAHLIGVHDDIVS 113
Db 276 AGVIRDRGVIPVDETMRTNPNYVIGDITGKLLAHVASHQGVIAKNISGHE---V 332
QY 114 MEYDLAYKGLDLPHTHTVISDIQDFVVALSLEISDEGNI--TMTSFEVRQ----- 161
Db 333 MDYSAIPSVITHP-----ETAMVGLSLOEAEQQNLPAKLTFFPKAIGKAVLGAS 384
QY 162 --FANVNNH-----IGLSILDPIFGVLSVLTAFQDTRKEMT 199
Db 385 DGFAAIVSHETIQILGAYVIGPHASSLIGEMTL----AIRNELT 425

RESULT 54
CZ02031
2-Oxo acid dehydrogenase, E3 component, lipamide dehydrogenase, probable CP1037 [imported]
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 03-Nov-2000
C:Accession: C72031; B81509
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72031
A:Molecule type: DNA
A:Residues: 1-461 <ARN>
A:CROSS-references: GB:AE001663; GB:AE001363; NID:g4377134; PIDN:AA018970.1; PID:g437714;
A:Experimental source: strain CML029
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, E.;
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: B81509
A:Molecule type: DNA
A:Residues: 1-461 <REA>
A:CROSS-references: GB:AE002261; GB:AE002161; NID:g7189950; PIDN:AAF38812.1; PID:g7189951
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: lpdA; CP1037
C:Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology
C:Keywords: redox-active disulfide
F:8-443/Domain: dihydrolipoamide dehydrogenase homology <DLD>
F:42-47/Disulfide bonds: redox-active #status predicted

Query Match 7.7%; Score 82; DB 2; Length 461;
Best Local Similarity 20.4%; Pred. No. 70;
Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;

QY 23 DKITEE-----INKALDDAIAIEQSETIDPMKVPDHDADKFE-----RH 61
Db 219 NKFTKQGRILTKA---SISAIESQNVRTVNDQVEEFYDLVAIGRQFNNTASIGLDN 275
QY 62 VGIV--DFKGELAMRNIEARGLKQMKRQD-----ANVKGEGIVKHAHLIGVHDDIVS 113
Db 276 AGVIRDRGVIPVDETMRTNPNYVIGDITGKLLAHVASHQGVIAKNISGHE---V 332
QY 114 MEYDLAYKGLDLPHTHTVISDIQDFVVALSLEISDEGNI--TMTSFEVRQ----- 161
Db 333 MDYSAIPSVITHP-----ETAMVGLSLOEAEQQNLPAKLTFFPKAIGKAVLGAS 384
QY 162 --FANVNNH-----IGLSILDPIFGVLSVLTAFQDTRKEMT 199
Db 385 DGFAAIVSHETIQILGAYVIGPHASSLIGEMTL----AIRNELT 425
```

RESULT 58
T38774
myosin-3 heavy chain, type II, myp2 [similarity] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 08-Sep-2000
C:Accession: T38774; T43276; T43553
R:Skelton, J.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Wood, V.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z21751
A:Accession: T38774
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2104 <SK>
A:Cross-references: EMBL:Z98762; PIDN:CAB11475.1; GSPDB:GN000066; SPDB:SPAC4A8.05C
A:Experimental source: strain 972h(-); cosmid c4A8

[illegible]

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Prohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsukano, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328

A:Contents: annotation
A:Genetics:
A:Gene: XF0395
A:Supertfamily: bacterioferritin

Query Match 7.6%; Score 81; DB 2; Length 154;
Best Local Similarity 21.7%; Pred. No. 22;
Matches 34; Conservative 23; Mismatches 42; Indels 58; Gaps 7;

QY 15 VSADPIHYDKITEINKAIDDAIAIEQ-----SETIDPMKVPD 53

Db 1 MKGDP---KVIEYLNKYLKELTAINQYFLHAKMLXNGLAEAEHEYESIDEMK--- 53

QY 54 HADKFERHVGIVD-----FKGELAMNIEARGLKQKMGQGDANVK 93

Db 54 HADRLSDRLFDGLPNFQALGKLRIGESPTIEFSGDALEQ-BAVAL---LREAISYTE 109

QY 94 GEEGIVKAHLIGV-----HDDIVSMEYDLAYKLG 124

Db 110 SVQDYVSRELLVDILNSEEHIDWLQDLIIKIGE 146

RESULT 62

R3MX4

ribosomal protein S4.eR - Methanococcus vannielii

N:Alternate names: ribosomal protein C

C:Species: Methanococcus vannielii

C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 22-Jun-1999

C:Accession: S05616

R:Auer, J.; Spicker, G.; Boeck, A.

J. Mol. Biol. 209, 21-36, 1989

A:Title: Organization and structure of the Methanococcus transcriptional unit homologous

S:ribosomes

A:Reference number: S05611; MUID:90040717; PMID:2530355

A:Accession: S05616

A:Molecule type: DNA

A:Residues: 1-244 <AUE>

A:Cross-references: EMBL:X16720; NID:944754; PIDN:CAA34692.1; PID:944760

C:Supertfamily: rat ribosomal protein S4

C:Keywords: protein biosynthesis; ribosome

Query Match 7.6%; Score 81; DB 1; Length 244;
Best Local Similarity 22.4%; Pred. No. 38;
Matches 36; Conservative 29; Mismatches 72; Indels 24; Gaps 6;

QY 17 ADPIHYDKITEINKAIDDAIAIEQSETIDPMKVPDHDKFERHVGIVDFKGLAMRNI 76

Db 56 ADNAREAKKIIONGLKILIDGVKREYKHPVGLMDVLIPELNENYLVLPENGRISLKT 115

QY 77 EARGLKQKMGQGDANVKGEIVKAHLIGVHD-----DIVSMEYDLAYKLGDLHP 127

Db 116 EKTGVKLCIKVNTKVIKG-----GHQLNLHGRNQIVKANALKAEDII-YKTGD--- 165

QY 128 TTHVISDIDQDFVVALSLEISDEGNITWTS--FEVRQFANVV 166

Db 166 --SVLVSLPEQAVVGVHVEF-NEGKLIYITGKHVGEFAKV 203

RESULT 63

F69119

cation efflux system protein (zinc/cadmium) - Methanobacterium thermoautotrophicum (stra

C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Feb-2003
C:Accession: F69119

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; F. ; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N. J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: F69119

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-298 <MTH>

A:Cross-references: GB:AE000941; GB:AE000666; NID:G2623025; PIDN:AAB96353.1; PID:G2623024

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1893

A:Start codon: TTG

Query Match 7.6%; Score 81; DB 2; Length 298;
Best Local Similarity 18.4%; Pred. No. 49;
Matches 41; Conservative 32; Mismatches 60; Indels 90; Gaps 9;

QY 77 EARGLKQKMGQGDANVKGEIVKAHLIGVHDIVSMEYDLAYKLGDL----- 125

Db 3 ESERIRLGKRAAFAGIGGNVLTSLNLFVGISSGSVALVAEAAHTLSVLTSTVITYIGFR 62

QY 126 -----HPTH-----VISDI-----QDFVVALSLE 145

Db 63 IQRPPDRQHPYHGGRALVGLVWVFLGIISYEILSEAYRKLFLELAPPDYTAALMAG 122

QY 146 ISDEGNITWTSFEVRQFANVV-----HIGLSILDLP 177

Db 123 FGIIANIAMTYY-IRRIGERINSPAIVADAQKQVDIFSCIAIMLGVAGSHL-GLRFDP 180

QY 178 IFGLSDVL---TALFQDTVRKEMTKVL---APAFKRELEKN 213

Db 181 LVAVIAIVLVLTAF--DVGRNVNIIILGAVPSPEIMEDIKS 221

RESULT 64

E70360

Glucose inhibited division protein A - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: E70360

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Owe

V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70360

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-445 <AQF>

A:Cross-references: GB:AE000702; NID:G2983276; PIDN:AAC06872.1; PID:G2983279; GB:AE000657

A:Experimental source: strain VF5

C:Genetics:

A:Gene: gidA2

Query Match 7.6%; Score 81; DB 2; Length 445;
Best Local Similarity 18.3%; Pred. No. 80;
Matches 55; Conservative 43; Mismatches 95; Indels 108; Gaps 11;

QY 13 VAVSADPIHYDKITEINKAID-----DAIAIEQSETIDPMK----- 50

Db 131 VVIATGPIITSALSEKIKELVGYDTLIFYDAIPIVEAESVDFSKGWSRYGKGGDDYF 190

QY 51 -----VPDHADKFERHV-----GIVDFKGEIAMRNIEARGLK-----QMKRQGD 89

Db 191 NCVLTEEBYKFKFYBELLKAERKPKDFKAVHFGCLPIIEEAERGYKTLFGPMKPYGL 250

QY 90 ANVK-GEIGIVKAHL-----LIGVHDDIVSMEYDLAY-----KLGD 125

hypothetical protein F14G11.4 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: F86386
R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.;
C.A.; Li, J.H.; Lister, J.Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F86386
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1201 <STO>
A:Cross-references: GB:AE005172; NID:gil1560181; PIDN:ARG38123.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 7.6%; Score 81; DB 2; Length 1201;
Best Local Similarity 25.4%; Pred. No. 2.7e+02;
Matches 47; Conservative 18; Mismatches 64; Indels 56; Gaps 9;

QY 19 PTHYDKITEINKAIDDAIAA-----IEQS-ETIDPMKVPDHADKFERHVGIV 65
 :
DB 779 PSFQVIRELDTKAGDAAATDLEVMEEPCIEGSVETEDPNPGSEADK-----T 831
 :
QY 66 DF-KGLAMRNTEARGLQMKRQGDANVKEGIVKAHLIGVHDDIVSMYEDVLAYKLGD 124
 :
DB 832 DIPKNDSDNAAAEAKKEKSPKKVKQ-----LVYELDD 869
 :
QY 125 LHP-----TTHVISIQDFVVALSLE-ISDSGNTMTSFV---RQFANVNHGGISIL 175
 :
DB 870 AHPHGFKAKTVLPD----VPAQQIEVVIRAGVSYNPLDVDPKIBELSNIIKGPSVI 925
 :
QY 176 DPIFG 180
 : : :
DB 926 HTLFG 930
 : : :

RESULT 67
B70451
prephenate dehydrogenase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 03-Mar-2003
C:Accession: B70451
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70451
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-311 <AQF>
A:Cross-references: GB:AE000754; NID:g2984047; PIDN:AAC07589.1; PID:g2984048; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: tyrA
C:Superfamily: prephenate dehydrogenase/arogenate dehydrogenase, feedback inhibition-inse

Query Match 7.5%; Score 80.5; DB 2; Length 311;
Best Local Similarity 22.9%; Pred. No. 56;
Matches 56; Conservative 34; Mismatches 88; Indels 67; Gaps 11;

QY 2 MKFLIIAAVFVAVS-ADPIH-----YDKITEINKAIDDAIAAEQSETIDPMK 50
 :
DB 30 MONVLIVGVGFMGGFSAKLSRSGFKGIYGYDINPESISKAVD--LCIIDEGTT----- 82

QY 51 VPDHAKFERHVGIVDFKGLAMENIARGIKOMKRO-----GDANVKGEEGIVKAHL 104
Db 83 -----STAKVEDSPDFVMJUSSPVRFREAKKLSVILSEDAIVT-DQSGVKGKLV 132
QY 105 IGVHDDIVSMEYDLAYKIGDLHPTTHVISDIQDFVVALSLEISDEGN---ITWTSFEVRQ 161
Db 133 -----YDENILKRFVGGHPAGTKEGSGVEYSLDNLVEGKKVILTPTKTKDK 181
QY 162 FANVNH-----IGGL-----SILDPTEGVLSD-----VLTAFQDQTVRKEMTKVL 202
Db 182 RLKLVXRVWEDVGGVMEYSPDLHDYVGVVSHLPHAVAFALVDLTHMSTPEVDLFKYP 241
QY 203 APAFK 207
Db 242 GGGFK 246
RESULT 68
C84040
hypothetical protein BH3123 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84040
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84040
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-374 <STO>
A:Cross-references: GB:AP001517; GB:BA000004; NID:gi0175500; PIN:BA06842.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3123
Query Match 7.5%; Score 80.5; DB 2; Length 374;
Best Local Similarity 22.4%; Pred. No. 71;
Matches 49; Conservative 35; Mismatches 74; Indels 61; Gaps 11;
QY 21 HYDKITEI---NKAIDDAIAAIQSETIDPMKVPDHDHAKFERHVGIVDFKGLAMENIE 77
Db 117 HLEKQEKELQKKEKELDEVGKERADKTIQPEK--DQLD-FEKRSANLD---LRQLQIQ 169
QY 78 ARGKQMKR-----QGDAVY-----KSEGVIVKAHLIGVHDDIVSM---EYDLA 119
Db 170 KGLEEQKDLVWKSRGDGTMTVNDHLQOGQDGMKPVLEIASAGELQLQKISEYD-- 227
QY 120 YKIGDLHPTTHVISDIQDFVVALSL--EISDEGNITWTSF-----EVRQFANV 165
Db 228 -----SLVISEEQVRISSDALPEEWEVGIVKESVYPPSEATLGNDAIGVQYPIIT 278
QY 166 VNHIGGLSILDPFQVLSLSDVLTAFQDQTVRKEMTKVLAP 204
Db 279 VQLSGDQSKLAPGYQMIMHIVT-----BEKTALTVP 309
RESULT 69
T47949
hypothetical protein F2A19.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47949
R:De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quest
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z24480
A:Accession: T47949
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-705 <DEH>
A:Cross-references: EMBL:AL132962
A:Experimental source: cultivar Columbia; BAC clone F2A19
C:Genetics:

A:Map position: 3
A:Introns: 61/3; 94/3; 135/3; 173/3; 210/1; 231/3; 287/3; 325/3; 372/3; 471/3; 493/3; 561
A:Note: F2A19.170
Query Match 7.5%; Score 80.5; DB 2; Length 705;
Best Local Similarity 24.6%; Pred. No. 1.5e+02;
Matches 41; Conservative 34; Mismatches 65; Indels 27; Gaps 9;
QY 27 BEINKAID-----DAIAAEQSETIDPMKV-PDHDHAKFERHVGIVD-----FK 68
Db 273 KEVNALEKENNELKKESELEALEESRKTKNSKVPFDATESTLTHSTLIDKEKPSFP 332
QY 69 G-----ELAMRNTEARGLKQKRGDANVKGBEGIVKAHLIGVHDDIVSMEYDLAYKIGD 124
Db 333 GKEMEQLQRLQ-MDLKETORERD-KARQELKRLKQHLLEKETESEKMDND-SRLIEE 389
QY 125 LHPTT-HVISDIQDFVVALSLEISD-EGNITWTSFEVRQFANVNH 169
Db 390 LRQTNQYORSQISHLEKSLKQAISSQEDNRLSNQIRKLDKDTVDL 436
RESULT 70
T45031
hypothetical protein Y39B68.e [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45031
R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton, J.
raser, A.; Fulton, L.; Gardner, A.; Green, P.; Hawkins, L.; Hillier, L.; Jier, M.; Johnston
B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopra, A.; Saunders, D.
Nature 369, 32-36, 1994
A:Authors: Showkeen, R.; Sims, M.; Smaldon, N.; Smith, A.; Smith, M.; Sonhammer, E.; St
tock, L.; Wilkinson-Sproat, J.; Wohlman, P.
A:Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.
A:Reference number: S43531; MUID:94150718; PMID:7906398
A:Accession: T45031
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1365 <WIL>
A:Cross-references: EMBL:AL132896; NID:G6434440; PIN:CA60910.1; PID:G6434445
A:Experimental source: clone Y39B68
C:Genetics:
A:Map position: 3
A:Introns: 10/1; 37/1; 171/2; 844/2; 1074/2; 1115/3; 1194/2; 1233/3
A:Note: Y39B68.e
Query Match 7.5%; Score 80.5; DB 2; Length 1365;
Best Local Similarity 21.5%; Pred. No. 3.5e+02;
Matches 43; Conservative 33; Mismatches 89; Indels 35; Gaps 8;
QY 27 BEINKAIDDAIAAEQSETID-PMKVPDHDHAKFERHVGIVDFKGL-----AM 73
Db 322 QSIKSLKPIFAFETLQAVNKKLVKVP-----FPHTKQFQKDEPYFIPVYNANSSVDVM 376
QY 74 RNIEARGLKQKRGDANVKGBEGI--VKAHLIGVHDDIVSMEYDLAYKIGDLHPTTHV 131
Db 377 EKYEGCDHEGRKYSDDYRSKGIEFINNVK-----IVSDHFTAFYFNFTQLDL 426
QY 132 ISDIQDFVVALSLEISDEGNITWTSFEVRQFANV--NHIGLSILDPFQVLSLDTAI 189
Db 427 ETDVAFVE--SLNTDINNRTQSEMEVIEIVKNSNNISLTKIQESVKGI-ENQFKAV 483
QY 190 FQDTRKEMTKVLAPAFKRE 209
Db 484 RLDSLKNDVFSKISDVFKD 503
RESULT 71
T27963
hypothetical protein ZK673.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Feb-2000
C:Accession: T27963

R;McMurray, A.
submitted to the EMBL Data Library, March 1995
A;Reference number: Z20447
A;Accession: T27963
A;Status: preliminary; translated from GB/EMBL
A;Molecule type: DNA
A;Residues: 1-160 <NIL>
A;Cross-references: EMBL:Z48585; PIDN:CAA88482
A;Experimental source: clone ZK673
C;Genetics:
A;Gene: CESP:ZK673.7
A;Map position: 2
A;Introns: 1/3; 39/3; 67/1; 107/3
C;Superfamily: calmodulin; calmodulin repeat h
C;Keywords: EF hand

Query Match	7.5%	Score 80;	DB 2;	Length 160;
Best Local Similarity	23.1%	Pred. No. 27;		
Matches	27;	Conservative 28;	Mismatches 42;	Indels 20;
Gaps				

Qy	53	DHADKPERHVGIVDFKGLAMR-----	NIEARGLKQMKRQGDANVKE-----	95
		: : : : : : : : : : : : :		
Db	15	DQIEQPKRYFNMPDKKGGYIRATQVGQILRTMGQAFERDLKQLIKEFDADGSGEIEFE	74	
		: : : : : : : : : : : : :		
Qy	96	--EGIVKAHLILGWDDIVSMEDYLAUKGLDLHPTHV--	ISDIQFVVALSLEISDE	149
		: : : : : : : : : : : : :		
Db	75	EFAAMVANFVNNENDEGLEELREAFRLYDKEGYINVSRLRIILRALDNVSEE	131	
		: : : : : : : : : : : : :		

RESULT 72
E72223
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C/Species: Thermotoga maritima
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C/Accession: E72223
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.

C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: E72223
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-292 <ARN>
A;Cross-references: GB:AE001809; GB:AE000512; NID:g4982257; PIDN:AAD36758.1; PID:g4982268
A;Experimental source: strain MSB
C;Genetics:
A;Gene: TM1691
C;Superfamily: hypothetical protein HI0457

Query Match	7.5%;	Score 80;	DB 2;	Length 292;
Best Local Similarity	18.6%;	Pred. No. 57;		
Matches	31;	Conservative 39;	Mismatches 53;	Indels 44; Gaps 6;
Qy	69	GELAMRNTEASGL-KQMKRQGDANVKGEGIVKAHLIGVHDDITVSMYDLAYKL-GDLH	126	
Db	40	GYSLSMKELELNNLVOEYVSRGKVQTRVQKVELEPPKVLDEIDKNVVVAYYSMLDSIVGELS	99	
Qy	127	-PTTHVIGDIOQFVVALSLEISDEGNITMTSFVRQRPANVNHIGGL-----	172	
Db	100	LPEPVKSLDLLNFREVFRIELSDE-----EIEINWNHVLVPILUREALEKXIVBERK	148	
Qy	173	-----SILDPI-----FGVLSDVLTALPQDTPVRKEMTKVL	202	
Db	149	KEGOKIGADLRKILLEDLSRVEEETEKISDQIPALYREKIKEEVPEKIL	195	

RESULT 73
F72326
hemolysin-related protein - Thermotoga maritima (strain MSB8)
C/Species: Thermotoga maritima
C/Date: 11-Jun-1999 #sequence revision 11-Jun-1999 #text chan

C/Accession: F72326
C/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A/Reference number: A72200; MUID:95287316; PMID:10360571
A/Accession: F72326
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-455 <ARN>
A/Cross-references: GB:AE001751; GB:AE000512; NID:G94981371; PIDN:AAD35927.1; PID:G94981371
A/Experimental source: strain MSB8
C/Genetics:
A/Gene: TM0845
C/Superfamily: hypothetical protein HI0107

Query Match	7.5%;	Score 80;	DB 2;	Length 455;
Best Local Similarity	19.1%;	Pred. No. 98;		
Matches 43;	Conservative 49;	Mismatches 79;	Indels 54;	Gaps 10
QY	25	ITEETNKALDAAIAEIOSETIDM-----KVPDHADKFERHVGIVDPFKGLAM-	73	
		: : : : : : : : : : : :		
Db	213	VREIMTPRD--IVAEEENQIVKDLIELVDEGYSRIPYVKETIDNIVGICVAKDVLMSL	270	
		: : : : : : : : : : : :		
QY	74	-----RNTEARGLKOMKQGQ-----DANVRGEGIVKA---HLLIGVHD-----DIVSME	115	
		: : : : : : : : : : : :		
Db	271	ABKDCEEVKSMMKVDIMREALYVPEWTNIDELIKLKARKIHIAIVVDEYGGTAGIVTLE	330	
		: : : : : : : : : : : :		
QY	116	-----YDLAYLGLDHPHTHVI---SDIQDFVALSLSEIDEGNITMTSP	157	
		: : : : : : : : : : : :		
Db	331	DIIEELFGNIMDEYDIDISGRKIDERTYIVDGNATPINDIEMELRVQPPETETETIAG	390	
		: : : : : : : : : : : :		
QY	158	EYRQFANVNVNHIGLSILDPFIIGVLSDLTAIFAQDVTYRKEMTKVL	202	
		: : : : : : : : : : : :		
Db	391	LEEKFRIPN-VGE-----EAVIGNLYFKVLAVGNRKETKWKIL	430	
		: : : : : : : : : : : :		

```

RESULT 74
S11210
probable
C:Species
C:Date: 2
C:Accessi
R:Jeffers
Nucleic A
A:Title:
A:Referen
A:Accessi
A:Molecul
A:Residue
A:Cross-r
C:Keyword

```

	Query Match	7.5%;	Score 80;	DB 2;	Length 798;
	Best Local Similarity	22.9%;	Pred. No. 2e+02;		
	Matches	36;	Conservative 29;	Mismatches 56;	Indels 36; Gaps 8
Qy	15	VSAADIHYDKITEEINKAIDDAIAAEISSETIPMKVPHAD-KFERHVGIV---DFKGE	70		
		: : :	: : :	:	:
Dd	574	VSAEKVNK---THSVNGITTEANTTIYSGKVIRPLRGVDPQTLEYOGMIEIVEEGDMKEG	630		
		: : :	: : :	:	:
Qy	71	LAMRNIEARGLKMQKRQGDANVKVEGGIKVAHLILIGVDHDIVSMEYDLA-----	119		
		: :	:	:	:
Dd	631	----VYPFGVGWANKGDCLOKGES--VFQLCV-LGNAQMATAINIPLRRATVECVC	682		
		: :	:	:	:
Qy	120	-----YKLGLDHPHTHHVISDIQDFVALSLSEISDE	149		
		: :	:	:	:
Dd	683	DOGFINTVEGDSKKLFFHFVKEVD---GHLEOAGDE	716		
		: :	:	:	:

RESULT 75
T19835

A; Cross-references: GB:AE005173; NID:G7715604; PIDN:AAPE68122.1; GSPDB:GN00141
C; Genetics:
A; Gene: F20B17.10
A; Map position: 1

Query Match 7.5%; Score 80; DB 2; Length 1487;
Best Local Similarity 22.9%; Pred. No. 4.2e+02;
Matches 52; Conservative 35; Mismatches 72; Indels 68; Gaps 13;

QY 33 IDDAIAAEQSETIDPMKVPDHADKPERHVGIVDFKGLAMPNINARGLKQMKRQG--- 88
DB 453 VDGRIIVAVKSKVWD-----EDKLEEFINEWILSQINHRNI-----VKLLGGCILE 498

QY 89 -DANVKGREGIVKAHLICGVHDDIVSMEY-----DLAYKLGDLH-----PTT 129
DB 499 TDVPLIVYEFINGNLFEHLHDD--SDDYTMTWEVRIRIADVIGALUSYLHSAASSIY 556

QY 130 HVISIDQDPVVALS-----LEISDEGNITMTSFEVRQFANVNHIGLSILDLP----- 177
DB 557 H--ROIKSTNIMLDEKHKRAKVSDFGSTRVTVDHHTLITVVS--GTVGYPMDPEYFQSSQF 612

QY 178 -----IFG-VLSDVLTIFAQDVT---RKEMTKVLAPAFKRELEKN 213
DB 613 TDKSDVSYFGVLABLITG--EKSVFGLRSQBYRTLATYFTLAMKEN 657

RESULT 77
F70062
hypothetical protein ywmB - Bacillus subtilis
C; Species: Bacillus subtilis
C; Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C; Accession: F70062
R; Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A; Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadiaie, Y.; Sato, T.; Scanlon,
A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T. T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A; title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A; Reference number: A93580; MUID:98044033; PMID:9384377
A; Accession: F70062
A; Status: preliminary; nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-246 <KUN>
A; Cross-references: GB:Z99122; GB:AL009126; NID:G2636029; PIDN:CAB15694.1; PID:G2636202
A; Experimental source: strain 168
C; Genetics:
A; Gene: ywmB
C; Superfamily: Bacillus subtilis hypothetical protein ywmB

Query Match 7.4%; Score 79.5; DB 2; Length 246;
Best Local Similarity 21.8%; Pred. No. 51;
Matches 47; Conservative 38; Mismatches 94; Indels 37; Gaps 9;

QY 5 LLIAAVAFVAVSADIHYDKITEINKAIDDAIAAEQSEITDPMKVPDHADKPERHVG 64
DB 11 IISVMLSFVIAVFHTIHASELF-----PLAQMAREGWERQDV--SIDKWTLHA-- 55

QY 65 VDFKGLAMRNTE---ARGLKQMKRQGDANVKGREGIVKAHLICGVHDD---IVSMEYD 117
DB 56 ---KQNLISLITEKFEYQKQRLKQFYQDVIAREDKMIKA---IGTYTDKKNRISFRLQ 109

QY 118 LAYKLGDLHPHTHVISDQDFVVALSLEISDEGNITMTSFEVRQFANVNHIGLSILD 177
DB 110 LVYTLKKNPTSLIYAE-----QMSLTPDSWNDTYEQFE--RETLGIFQE--KVVIETC 160

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A'Reference number: A99758; MUID:21311952; PMID:11418146
A'Accession: G9894
A>Status: preliminary
A'Molecule type: DNA
A'Residuals: 1-664 <KUR>
A'CROSS-references: GB:BA000018; PID:g13701020; PIDN:BA42315.1; GSPDB:GN00149
A'Experimental source: strain N315
C'Genetics:
A'Gene: SA1063

Query Match 7.4%; Score 79.5; DB 2; Length 664;
Best Local Similarity 21.6%; Pred. No. 1.7e+02;
Matches 35; Conservative 31; Mismatches 51; Indels 45; Gaps 5;
QY 1 MKKELLIAAVFVAVSADPHYDKITEINKAIDDAIAEIOSETIDPMKV----- 51
DB 355 IFSLLMIALVSVAMFGNKYEETPDVIGSVKEA-----EQIFNKNLKLGIKISRSYSD 410
QY 52 -----PDHAKFER--HVGIVDFK-----ELAMRNIEARGLQK 84
DB 411 KYPENELIKTPTGTERVERGSDVDVSKGPKVKNVIGLKPKEALQKLSGLKDV 470
QY 85 KROGDANVKGEG-----VKAHLIGVHDDIVSMYVDLAYK 121
DB 471 TIEKYNNQAPKGYANOSVTAETIAIHDSNKLKYESLGK 512

RESULT 82
D90541
acp synthase beta chain [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C'Species: Mycoplasma pulmonis
C'Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C'Accession: D90541
R'Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallissou, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A>Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*.
A'Reference number: A99512; MUID:21267165; PMID:11353084
A'Accession: D90541
A>Status: preliminary
A'Molecule type: DNA
A'Residuals: 1-698 <KUR>
A'CROSS-references: GB:AL445566; PID:g14089649; PIDN:CAC13409.1; GSPDB:GN00153
A'Experimental source: strain UAB CTIP
C'Genetics:
A'Gene: MYPU 2360
A'Genetic code: SGC3

Query Match 7.4%; Score 79.5; DB 2; Length 698;
Best Local Similarity 23.0%; Pred. No. 1.8e+02;
Matches 46; Conservative 31; Mismatches 80; Indels 43; Gaps 7;
QY 24 KITEINKAIDDAIAEIOSETIDPMKVPHADKFERHVGIVDFKGLANR---NIEARG 80
DB 155 KIFPENKTKQSLIAAIAERQ-----YTDDFR-----KSEIMKPRAKFKARN 198
QY 81 -LKQMKROGDANVKGEGIVKAHLIGVHDDIVSMYVDLAYKGLDHPHTHVISDIO-DF 138
DB 199 PLKPLED-----KAYVLEGIVVEKQVVKIKLTVKEPIINALFEIQTEQ 245
QY 139 VVALSLEISDEGNITWTS-----FEVRQFANVNVHIGGSLIDPIFGVLSDLTAT 189
DB 246 QOTRLLLEISLSDSLVAGYVGLGREGIETGFSFARKSNPNPYSIPSEKLLRIIDPVGRI 305
QY 190 FQDITVRKMTKVLAPAPKRE 209
DB 306 LDDPHTPLVGKQYAPMLETE 325

RESULT 83

C71527
endopeptidase La (EC 3.4.21.53) - *Chlamydia trachomatis* (serotype D, strain UM3/Cx)
N'Alternate names: ATP-dependent proteinase Lon; ATP-dependent serine proteinase La
N'Contains: adenosinetriphosphatase (EC 3.6.1.3)
C'Species: *Chlamydia trachomatis*
C'Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Jun-2003
C'Accession: C71527
R'Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, S.; Science 282, 754-759, 1998
A'Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*.
A'Reference number: A71570; MUID:9900809; PMID:9784136
A'Accession: C71527
A>Status: preliminary
A'Molecule type: DNA
A'Residuals: 1-819 <ARN>
A'CROSS-references: GB:AE001307; GB:AE001273; NID:g3328757; PIDN:AAC67939.1; PID:g3328764
A'Experimental source: serotype D, strain UM-3/Cx
C'Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP
C'Genetics:
A'Gene: lon
C'Superfamily: ATP-dependent Lon protease
C'Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine I
F;392-399/Region: nucleotide-binding motif A (P-loop)
F;455-460/Region: nucleotide-binding motif B
F;724/Active site: Ser #status predicted

Query Match 7.4%; Score 79.5; DB 1; Length 819;
Best Local Similarity 21.1%; Pred. No. 2.2e+02;
Matches 51; Conservative 35; Mismatches 73; Indels 83; Gaps 10;

QY 16 SADPIHYDKITEINKAIDDAIAEIOSETIDPMKVPH----- 54
DB 6 NTDSQNLDPNASEVEKLLDESAAEE-----KVDDHTPPSELPILNKRPFPGMA 57
QY 55 -----ADKFERHVGIVDFKGLANRNIARGLQKMKROGDANVKGEG 97
DB 58 APLLIEAGPHYEVLLTAKSKQKHIGLVLTCKEDA--NTLVGFNQLHRVG----- 106
QY 98 IVKAHLIGVHDDIVSMYVDLAYKGLDHPHTHVISDIOQDFWALSLEISDEGNITWTSF 157
DB 107 -VSARIL-----RINPIEGGSAQVLLSIEDRIRIVKPIQDKYLRKAKVSYHKE-NKELTE- 158
QY 158 EVRQFA-NVNVHIGGSLIDPIF-----GVLSDLTATFQDITVRKMTK 200
DB 159 ELKAYSISIVSIKDLKLNPLFKEELQIFLGHSDFTPEFKLADFSVALTTAT-REELQE 217
QY 201 VL 202
DB 218 IL 219

RESULT 84
A43000
alpha N-catenin - chicken
C'Species: Gallus gallus (chicken)
C'Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C'Accession: A43000
R'Hirano, S.; Kimoto, N.; Shimoyama, Y.; Hirobaishi, S.; Takeichi, M.
Cell 70, 293-301, 1992
A'Title: Identification of a neural alpha-catenin as a key regulator of cadherin function.
A'Reference number: A43000; MUID:92346716; PMID:1638632
A'Accession: A43000
A>Status: preliminary
A'Molecule type: mRNA
A'Residuals: 1-906 <HIR>
A'CROSS-references: GB:D11090; NID:g222787; PIDN:BAA01863.1; PID:g222788
A'Experimental source: embryonic brain
A>Note: sequence extracted from NCBI backbone (NCBI:109985, NCBI:109986)
C'Superfamily: alpha-catenin; vinculin amino-terminal homology; vinculin carboxyl-terminal
C'Keywords: cytoskeleton
F;19-267/Domain: vinculin amino-terminal homology <VINN>
F;375-859/Domain: vinculin carboxyl-terminal homology <VINC>

Db 135 NDHIFNPEQHRIVTAASCTTNCIAPVVKVIEHKLGAQASFTTTHNLNTOTILDAPHKD 194
QY 142 -----LSLEISDEGNITWTSFEVQCPANVNVHIGGLSILDPICGV-LSVDVLTALFQD 192
Db 195 LRRARACMSLIPTTTSAKAI-----IEIFPLKQKIDGHAVRVPLANASUTDIIFDVQD 251
QY 193 TVRKEMTKVLAPAFKRELE 211
Db 252 TTVEEINQLLKQASENELK 270
RESULT 88
G90203
ethylene-inducible protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: G90203
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90203
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <KUR>
A:Cross-references: GB:AB006641; NID:gl13813736; PIDN:AAK40886.1; GSPDB:GN00155
C:Genetics:
A:Gene: SS00570
C:Superfamily: hypothetical protein H11647
Query Match 7.4%; Score 79; DB 2; Length 338;
Best Local Similarity 24.8%; Pred. No. 82;
Matches 37; Conservative 22; Mismatches 46; Indels 44; Gaps 7;
QY 13 VAVSADPIHYDKITEINKAIDDAIAA-----IEQSEITDPMKVP 52
Db 100 VARWADP-----KITEVNSSITIPVMAKVIGHYVEAKLEALGVMDISEVLTADBE 155
QY 53 DHADKFRHVGIVDFK---GELAVRNIEARGLKQKQGDANVKGEEGIVKAHLIGVHD 109
Db 156 HHINKWFSPVFNAGRNIGALRRTSE--GASMRKTGEAGTGNVSEAVK-HMKI-INS 211
QY 110 DIVSM-----EYDLAYKLGLD 125
Db 212 EIRLSISMSEEDRVKAREYQVPYQIVEL 240
RESULT 89
AI3312
ATP synthase beta chain/transcription termination factor rho BMEI0487 [imported] - Brucel
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AI3312
R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:1175668
A:Accession: AI3312
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51568.1; PID:gl17982399; GSPDB:GN00190
C:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0487
A:Map position: I
Query Match 7.4%; Score 79; DB 2; Length 401;
Best Local Similarity 25.5%; Pred. No. 1e+02;
Matches 59; Conservative 20; Mismatches 82; Indels 70; Gaps 13;

QY 12 FVAVSADPIHYDKI--TEEINKAI-----DDAIAAIEQSE--TIDP 48
Db 180 FADIADTVHGLIILVTNELFDALPFRQFVADGRFVERMIALNEQDFQFVSGAGGIDP 239
QY 49 MKVP--DHADK-----FERHVGIVDFKGLAMRNIEARGLKQKQGDANVKGEEGIVKAH 102
Db 240 ALLPKDHVKAEGALFEAAPARTALMQBIASRIATRG-----AALNIDYGH 286
QY 103 LLIGVHDDIVSM--EYDLAYKLGLDHP-----TTHVISDI-QDFVVALSLEISDEGNIT 153
Db 287 LESGFGDTLQAMLKQAYDDVF-----AHGVSADLTSHVDLDTAKACGCKTG-----T 337
QY 154 MTSFEVQCPANVNVHIGGLSILDPICGVLSVDVLTALFQDTRKEMTKVLP 204
Db 338 MTQGEFLLAMGLVDRAGRL-----GAGKD--AAFOEKIQDVERLAAP 378
RESULT 90
AC2152
dolichyl-phosphate-mannose synthase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2152
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2152
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074469.1; PID:gl17131863; GSPDB:GN00179
C:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2770
C:Superfamily: stress response protein csbB
Query Match 7.4%; Score 79; DB 2; Length 412;
Best Local Similarity 21.0%; Pred. No. 1e+02;
Matches 41; Conservative 36; Mismatches 74; Indels 44; Gaps 8;
QY 15 VSADPIHYD-----KITEINKAIDDAIAAIEQSEITDPMKVPDHADKFERH 61
Db 29 VTGQPIQFSLIPIPTVKESGNTNRNVIERSQLDDEIFGDIYELIVDD-DSPD--GTWEVA 85
QY 62 VGI-VDFKGLAMRNIEARGLKQKQGDANVKGEEGIVKAHLIGVHDIVSMYDLAY 120
Db 86 LSLTAEPQLRVMRQBERGLSTAVIRGQVARGS-----ILGVID----- 126
QY 121 KLGLDHPHTHVIDIQDFV-----VALSLEISDEGNITWTSFEVQCPANVNVHIGGLSIL 175
Db 127 --GDLQHPHYLLLELLSKHKGADLAVASRHVDGGVGSWSF-IRRFSLRGAQLLGLVIL 183
QY 176 DPFGVLSVDVLTALF 190
Db 184 PSVLGRVSDPMGYSYF 198
RESULT 91
A75045
2-isopropylmalate synthase (leua-1) PAB0890 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: A75045
R:Anonymous, Genoscape
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: A75045
A>Status: preliminary

[illegible]

Qy 114 MEYDLAYKGLDLPHTTHV--ISDIQDFVVALSL---BISDEGNITWT 155
Db 111 VT-----NYAPSEMIKOISDVNIEFVAISLTGEVGEKGLNPT 149

RESULT 95

D69064
isopropylmalate synthase - Methanobacterium thermoautotrophicum (strain Delta H)
C/Species: Methanobacterium thermoautotrophicum
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
C/Accession: D69064
R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.;
ki, S.; Church, G.M.; Daniels, C.G.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A/Reference number: A69000; MUID:98037514; PMID:9371463
A/Accession: D69064
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-505 <MTH>
A/Cross-references: GB:AE000909; GB:AE000666; MID:g2622596; PIDN:AAB85956.1; PID:g2622596
A/Experimental source: strain Delta H
C/Genetics:
A/Gene: MTH1481
C/Superfamily: 2-isopropylmalate synthase leuA

Query Match 7.4%; Score 78.5; DB 2; Length 505;
Best Local Similarity 20.6%; Pred. No. 1.5e+02;
Matches 40; Conservative 36; Mismatches 73; Indels 45; Gaps 8;

Qy 30 NKAI--DDAIA-----AIEQSEITDPM--KVPDHADK--FERHVGIVDFKGLAMR 74
Db 284 NKAIVGENAFAHEAGIHVHGVLKAEATYEIPITMVGHKRIVLGKXTGNALRSKLQEV 343
Qy 75 NIBARG-----LKQMKRQGDANV-----KGEEGIVKAHLFIGVHDDIV 112
Db 344 GIMKEBEQFCTLYEQVKRLGDKGRITDADLRAMAVTILGKASREIVKLEGIAVMTGESV 403
Qy 113 SMEYDLAYKGLDLPHTTHV--ISDIQDFVVALSLSDGNITMTSFEVYRQFANVNHIG 170
Db 404 MPATVKLRIGDEVKTSMTGCVGPVDAINAIQSLVSETADIELDSEYIEAT-----G 457
Qy 171 GLSILDPPIQVLSD 184
Db 458 GTNALAEVFMVMSD 471

RESULT 96

S70295
probable membrane protein YAL028w - yeast (Saccharomyces cerevisiae)
C/Species: Saccharomyces cerevisiae
C/Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 19-Apr-2002
C/Accession: S70295; S51993; S51992
R/Vo, D.
submitted to the EMBL Data Library, April 1996
A/Reference number: S70291
A/Accession: S70295
A/Molecule type: DNA
A/Residues: 1-528 <VOD>
A/Cross-references: EMBL:U12980; MID:g1326053; PID:g1326061; MIPS:YAL028w
R/Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Quel
submitted to the EMBL Data Library, August 1994
A/Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A/Reference number: S51956
A/Accession: S51993
A/Molecule type: DNA
A/Residues: 167-528 <BUS>
A/Cross-references: EMBL:U12980; MIPS:YAL028w
A/Note: this sequence has been revised in reference S70291
C/Genetics:
A/Cross-references: SGD:S0000026
A/Map position: 1L

C;Keywords: transmembrane protein
F;505-521/Domain: transmembrane #status predicted <TM1>

Query Match 7.4%; Score 78.5; DB 2; Length 528;
Best Local Similarity 23.0%; Pred. No. 1.5e+02;
Matches 41; Conservative 27; Mismatches 65; Indels 45; Gaps 8;

Qy 39 AIEQSTIDPMKVP-----DHADKFERHVGIVDFKGELAVRNIEARGLQM 84
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 328 SFQSVAVDPLPPGNTSYSSNLSINSDLYVQRHG-----LQLQQTALLKEHS 380
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 85 KRGDANVAGSGIYKAHLIGVDDHDIYSMEYDLAYKLGDLPHTTHVISDI--QDFVVAL 142
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 381 K---DSVLKDNDLVKN---IANFDKIVK-----EURLRSRTIGWKEVEEDYLNNL 427
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 143 SLEISDEGNITTSFEVRQFANVANVHIGLSILDPIFGVLSDLVLTAFIQDTVRKEMTK 200
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 428 QKDFDKR--NPESPEARLSDTINTVAKIQDLEKRMSCKDRLAS-----RKEVMR 476

RESULT 97
AB2401
hypothetical protein all4762 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2401
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Triguichi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana-
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2401
A>Status: preliminary
A:Molecule type: DNA
A;Residues: 1-573 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76461.1; PID:g17133899; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all4762

Query Match 7.4%; Score 78.5; DB 2; Length 573;
Best Local Similarity 24.1%; Pred. NO. 1.7e+02;
Matches 51; Conservative 36; Mismatches 84; Indels 41; Gaps 10;

Qy 23 DKITEINKAIDDATAAEQSETID---PMKVPHADKF--ERHVGIVDFKGLAMVN- 75
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 358 DLALAEID-AIFARLENALKLRQENQILENQOEDSLNKIAHEVHKNIKTRSLAQEND 416
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 76 IEARGLKMKRGQDANKVGEIGIVKAHLIGVDDHDIYSMEYDLAYKLGDLPHTTHVISDI 135
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 417 AVAQCMEVWKGLANLNKQGLDYFL---HNEAFSL-LELAADSQRFLLIQADAVSPI 471
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 136 -----QDFVVALSLEISDEGNITMTSFVRQFANVVNHIGLSILD----- 176
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 472 PNQPEWKIRCODFILLUS-GVSDDIRIGATFWLPRPFLSUTLLV-GLSLVGISGLYYDN 529
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Qy 177 -----PIFGVLSDLVLTAFIQDTVRKEMTKV 201
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :
Db 530 GLTFGARFPFDYLGILWGISADVASRLSNL 561
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :

RESULT 98
Tl3646
hypothetical protein 95B7.9 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: Tl3646
R;Ferraz, C.; Vidal, S.; Brun, C.; Bucheton, A.; Demaille, J.G.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17694
A;Accession: Tl3646

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-735 <PER>
A/Cross-references: EMBL:AL021728; PIDN:CAAL6821.1
C:Genetics:
A/Cross-references: FlyBase:FBgn0000376
A:Introns: 39/3; 93/2; 294/2; 387/2
A:Note: EG:95B7.9
C:Superfamily: fruit fly hypothetical protein EG_95B7.9

Query Match 7.4%; Score 78.5; DB 2; Length 735;
Best Local Similarity 24.7%; Pred. No. 2.3e+02;
Matches 41; Conservative 21; Mismatches 57; Indels 47; Gaps 8;
QY 20 IHDKITEEINKAIDD-----AIAIEQSETIDPMKVPDHADKPERHVGIVDF----- 67
Db IRYDNSQVIAAILDDNLPDLAQMDRQEVIP---PDQDKQQRQTGLRHFNVHGDY 534
QY 68 -----KG-ELAMRNE-----ARGLKQMKRQGDANVKGEGIVKAHLIG 106
Db 535 DVLTRDQPECIKQKGLPGAPRNAEQLLDDKDLKQKERYQKYAMVEETTLER-----G 590
QY 107 VHDIVSMEYDLAYKL--GDLHPTTHVISDIQDFVVALSLEISDE 149
Db 591 EYDD---EYDSDVEAQNEQAPVSLIRGLQEAASSAYNAQDE 632

RESULT 99

F89819
endopeptidase [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 17-May-2002
C:Accession: F89819
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mitutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsesu, K. Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89819
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-818 <KUR>
A/Cross-references: GB:BA000018; PID:gl3700415; PIDN:BAB41713.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: clpC
C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 7.4%; Score 78.5; DB 2; Length 818;
Best Local Similarity 22.6%; Pred. No. 2.6e+02;
Matches 49; Conservative 37; Mismatches 86; Indels 45; Gaps 11;
QY 18 DPHYDKITEE-----INKAIDDAIAIEQSETIDPMKVPDHADKPERHVGIVDFKGE 72
Db 504 DTLHERVIGQDAVNSISKAVRRARAGLK-----DP-----KRPFGSIFLGLPTG 548
QY 73 MNIE-ARGL-KQMKRQGDANVKG-EGIVKAHL--LIGV-----HDDIVSMEYDLAY 120
Db 549 VGTALARALAESMFGDDAMIRVDMSEFMEKHAIVSLVAPPYVGHDDGGQTEKVR 608
QY 121 KLGDLHPTTHVISD-----IQDFVVALSLEISDEGNITMTSFEVQFANVV-----NHIGGL 172
Db 609 K-----PYSVILFDEIEKAHPDVFNILLQVLDDGHLTDTKGRTVDFRNTIIIMTSNVGAQ 663
QY 173 SILDPIFGVLSDLTAIFQDTRKEMTKVLAPFRE 209
Db 664 ELQDQRFAGFGGSSDQDYETIRKTMKLNKNSRPE 700

RESULT 100

E81681
proteinase, Lon family TC0623 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2003
C:Accession: E81681
R;Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, I.; C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: E81681
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <FEM>
A/Cross-references: GB:AE002331; GB:AE002160; NID:g7190662; PIDN:AAF39454.1; PID:g7190662
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0623
C:Superfamily: ATP-dependent Lon protease

Query Match 7.4%; Score 78.5; DB 2; Length 819;
Best Local Similarity 21.8%; Pred. No. 2.6e+02;
Matches 51; Conservative 38; Mismatches 78; Indels 67; Gaps 11;
QY 16 SADPHYDKITEEINKAIDDAIAIEQSETIDP-----MKVP----- 52
Db 6 NTDSONLPNASEVEKLLIDESAEEKTDDHTPPSELFILPLNKPPFPFGMAAPLLIEAG 65
QY 53 DH-----ADKFERHVGIVDFKGLAMRNEARGLKQMKRQGDANVKGEGIVKAHLI 105
Db 66 PHVEVTLIAKSSQKHIGIVLTKEKA--NTLKIGFNQLHRVG-----VSARIL-- 112
QY 106 GVHDDIVSMEYDLAYKLGDLHPTTHVISDIQDFVVALSLEISDEGNITMTSFEVQFAN 164
Db 113 ----RIMPEGGAQVLLSIEDRIIVKPVQDKYKAKVAHKE-NKELTE-ELKAYSIS 166
QY 165 VVNHIGGLSILDPF-----GVLSDLTAIFQDTRKEMTKVL 202
Db 167 IVSIIKDLKLNPLFKEELQIFLGHSDFTEPGKLADFSVALTTAT-RBELQEV 219

Search completed: August 6, 2004, 16:02:10
Job time : 20 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 15:54:41 ; Search time 14 Seconds
(without alignments)
792.210 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068

Sequence: 1 MMKFLIIAAVAVVNSADPI.....VRKMTKVLAPAFKRELEKN 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1068	100.0	213	1	ALL7_DERFA
2	962	90.1	215	1	ALL7_DERPT
3	250.5	23.5	216	1	ALL7_LEPDS
4	103	9.6	436	1	ENGA_STRMU
5	91.5	8.6	1393	1	RPOC_CHLPN
6	90.5	8.5	547	1	CH60_XYLFA
7	90.5	8.5	547	1	CH60_XYLFT
8	90.5	8.5	547	1	CH61_VIBFA
9	89.5	8.4	1539	1	Y373_HUMAN
10	88.5	8.3	458	1	FKB4_HUMAN
11	88.5	8.3	1453	1	Y373_BOVIN
12	88	8.2	268	1	SPED_BUCAP
13	88	8.2	436	1	ENGA_STRPN
14	88	8.2	547	1	CH60_LEGPN
15	87.5	8.2	1011	1	SCA4_RICMN
16	87	8.1	436	1	ENGA_STRP3
17	87	8.1	436	1	ENGA_STRP5
18	86.5	8.1	457	1	FKB4_MOUSE
19	86.5	8.1	544	1	CH61_VIBCT
20	85.5	8.0	553	1	CH60_BUCTH
21	85.5	8.0	903	1	YB56_MERJA
22	85.5	8.0	1396	1	RPOC_CHLMU
23	85.5	8.0	1396	1	RPOC_CHLTR
24	85	8.0	550	1	CH60_BUCTS
25	84.5	7.9	539	1	CH60_SERRU
26	84.5	7.9	540	1	CH60_KLEOR
27	84.5	7.9	540	1	CH60_KLEOX
28	84.5	7.9	1887	1	RPB1_DROME
29	84	7.9	358	1	SYFA_STRCO
30	84	7.9	392	1	DEOB_STAAM
31	84	7.9	513	1	XYLG_ECOLI
32	83.5	7.8	540	1	CH60_ENTGE
33	83.5	7.8	546	1	CH61_VIBVU

RESULT 1

ALIGNMENTS

34	83.5	7.8	548	1	CH60_YERPE
35	83.5	7.8	550	1	CH60_YEREN
36	83	7.8	244	1	R34E_METJA
37	83	7.8	244	1	APBA_CLOAB
38	83	7.8	465	1	PUR9_SYNEL
39	83	7.8	518	1	CH60_BACFO
40	83	7.8	543	1	DNAK_AGRTO
41	83	7.8	633	1	CH60_AGRTO
42	82.5	7.7	1060	1	DF3A_LACLA
43	82.5	7.7	279	1	CH60_METJA
44	82.5	7.7	539	1	CH60_ENTAM
45	82.5	7.7	540	1	CH60_KLEPL
46	82.5	7.7	540	1	CH60_SERFI
47	82.5	7.7	546	1	CH60_XANAC
48	82	7.7	2483	1	COA2_HUMAN
49	82	7.7	461	1	DLDH_CHLPN
50	81.5	7.6	2104	1	MY53_SCHPO
51	81.5	7.6	541	1	CH60_PANAN
52	81.5	7.6	547	1	CH60_HAEDU
53	81.5	7.6	548	1	CH61_VIBHA
54	81	7.6	745	1	ATCS_SYNY3
55	81	7.6	244	1	R34E_METJA
56	81	7.6	445	1	GID_AQUAE
57	80.5	7.5	159	1	UNR_HUMAN
58	80.5	7.5	194	1	VATE_SULTO
59	80.5	7.5	539	1	CH60_ENTAG
60	80.5	7.5	539	1	CH60_ENTAS
61	80.5	7.5	540	1	CH60_ENTAE
62	80.5	7.5	540	1	CH60_ERWAP
63	80.5	7.5	541	1	CH60_ERWHE
64	80.5	7.5	1010	1	SCA4_RICPA
65	80	7.5	160	1	TPC2_CABEL
66	80	7.5	590	1	SYFB_METJA
67	80	7.5	798	1	UNR_RAT
68	80	7.5	1816	1	LMA4_MOUSE
69	79.5	7.4	457	1	FXB4_RABIT
70	79.5	7.4	540	1	CH60_ERWCA
71	79.5	7.4	547	1	CH60_SALTY
72	79.5	7.4	819	1	CON_CHLTR
73	79.5	7.4	906	1	CTN2_CHICK
74	79	7.4	153	1	ARGR_CLOFE
75	79	7.4	220	1	KPYK_THELI
76	79	7.4	406	1	RL4B_ARATH
77	79	7.4	486	1	LE11_PYRAB
78	79	7.4	867	1	SYA_FUSNN
79	78.5	7.4	326	1	LIPC_SOLTU
80	78.5	7.4	351	1	YB72_HAETN
81	78.5	7.4	505	1	LEA2_METHR
82	78.5	7.4	528	1	YAC8_YEAST
83	78.5	7.4	532	1	CH62_VIBVU
84	78.5	7.4	546	1	CH60_XANCH
85	78.5	7.4	549	1	CH60_COLMA
86	78.5	7.4	819	1	CON_CHLMU
87	78.5	7.4	892	1	SYA_METJA
88	78.5	7.4	1081	1	SYT_TETTH
89	78	7.3	328	1	SYFA_BUCBP
90	78	7.3	407	1	RL4A_ARATH
91	78	7.3	555	1	CH60_RICTS
92	78	7.3	664	1	Y4FB_RHISN
93	78	7.3	1012	1	POLS_IBDV5
94	78	7.3	1012	1	POLS_IBDV5
95	77.5	7.3	191	1	VATE_SULAC
96	77.5	7.3	458	1	RADA_BACSU
97	77.5	7.3	539	1	CH60_ENTIT
98	77.5	7.3	545	1	CH60_SODGL
99	77.5	7.3	546	1	CH60_ACTPL
100	77.5	7.3	546	1	CH60_XANCP

Q8ziy3	yersinia pe
P48219	yersinia en
P54039	methanococc
Q97k30	clostridium
Q8dins	s bifunctio
P81284	bacteroides
P50019	agrobacteri
Q9c170	lactococcus
Q58488	methanococc
O66196	enterobacte
O66212	klebsiella
O66204	serratia fi
O8ppz1	xanthomonas
O00763	homo sapien
Q92773	chlamydia p
O41517	schizosacch
O66218	pantoea ana
P31294	haemophilus
Q83wi8	vibrio harv
P73241	synechocyst
P40233	methanococc
O66913	aquifex aeo
O75534	homo sapien
O68935	serratia ma
Q971b8	sulfolobus
O66200	enterobacte
O66198	enterobacte
O66222	erwinia aph
O66216	erwinia her
Q9aj75	rickettsia
O09665	caenorhabdi
Q8tx56	methanopyru
P13395	rattus norv
P79727	mus musculu
P27124	oryctolagus
O66220	erwinia car
P48217	salmonella
O84348	chlamydia t
P30997	gallus gall
Q894h3	clostridium
Q86301	thermococcu
Q98f40	arabidopsis
Q9uz08	pyrococcus
Q8xfj8	fusobacteri
P80471	solanum tub
P44206	haemophilus
O27525	methanobact
P39734	saccharomyc
Q8cwj0	vibrio vuln
Q8xit7	xanthomonas
Q93gt8	colwellia m
Q9pk50	chlamydia m
O57984	methanococc
P36422	tetrahymena
P59504	buchnera ap
P49691	arabidopsis
P16625	rickettsia
P55440	rhizobium s
P25219	avian infec
P23511	avian infec
P22722	sulfolobus
P37572	bacillus su
O66192	enterobacte
Q9anr8	sodalis glo
P94166	actinobacil
Q8pd23	xanthomonas

ALL7 DERFA
ID ALL7_DERFA STANDARD; PRT; 213 AA.
AC Q26456;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mite allergen Der f 7 precursor (Der f VII).
GN DERF7.
OS Dermatophagoides farinae (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Psoroptidia; Analgoidea;
OC Pyroglyphidae; Dermatophagoides.
OX NCBI_TaxID=6954;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120794; PubMed=8556554;
RA Shen H.-D., Chua K.-Y., Lin W.L., Hsieh K.-H., Thomas W.R.;
RT "Molecular cloning and immunological characterization of the house
RT dust mite allergen Der f 7."
RL Clin. Exp. Allergy 25:1000-1006(1995).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALLERGEN: Causes an allergic reaction in human.
CC -!- SIMILARITY: Belongs to the mite group 7 allergen family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U37044; AAA80264.1; -.
KW Allergen; Signal.
FT SIGNAL 1 17
FT CHAIN 18 213
FT CARBOHYD 151 151
FT SEQUENCE 213 AA; 23627 MW; 3CF1F529107B7808 CRC64;
SQ
Query Match 100.0%; Score 1068; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 3.4e-79;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDADKFER 60
DB 1 MKKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDADKFER 60
QY 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAY 120
QY 121 KGLDLHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFG 180
DB 121 KGLDLHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFG 180
QY 161 VLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEKN 213
DB 161 VLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEKN 213
QY 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAY 120
QY 121 KGLDLHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFG 180
DB 121 KGLDLHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFG 180
QY 161 VLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEKN 213
DB 161 VLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEKN 213
RESULT 2
ALL7 DERPT
ID ALL7_DERPT STANDARD; PRT; 215 AA.
AC P49273;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mite allergen Der p 7 precursor (Der p VII).
GN DERP7.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Psoroptidia; Analgoidea;
OC Pyroglyphidae; Dermatophagoides.

OX NCBI_TaxID=6956;
RN [1]
RP SEQUENCE FROM N.A.
RX Shen H.-D., Chua K.
RT "Molecular cloning
RT binding specificit
CLin. Exp. Allergy
CC -!- SUBCELLULAR LO
CC -!- ALLERGEN: Caus
CC -!- SIMILARITY: Be
CC This SWISS-PROT en
CC between the Swiss
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U37044; AAA80264.1; -.
KW Allergen; Signal.
FT SIGNAL 1 17
FT CHAIN 18 215
FT CARBOHYD 151 151
FT SEQUENCE 215 AA; 23881 MW; 63AD03DB5C8B1C0 CRC64;
SQ
Query Match 90.1%; Score 962; DB 1; Length 215;
Best Local Similarity 85.9%; Pred. No. 1.2e-70;
Matches 193; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDADKFER 60
DB 1 MKKFLIIAAVAVFVAVSADPIHYDKITEINKAIDDAIAAIQSETIDPMKVPDHDADKFER 60
QY 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAY 120
DB 61 HVGIVDFKGLAMRNIEARGIKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMYDLAY 120
QY 121 KGLDLHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFG 180
DB 121 KGLDLHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPIFG 180
QY 161 VLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEKN 213
DB 161 VLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEKN 213
RESULT 3
ALL7 LEPDS
ID ALL7_LEPDS STANDARD; PRT; 216 AA.
AC Q9ULG2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mite allergen Lep d 7 precursor.
OS Lepidoglyphus destructor (Storage mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Glycyphagoidea; Glycyphagidae;
OC Lepidoglyphus.
OX NCBI_TaxID=36936;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2109292; PubMed=11168362;
RA Eriksson T.L.J., Rasool O., Huecas S., Whitley P., Cramer R.,
RA Appenzeller U., Gafvelin G., van Hage-Hamsten M.;
RT "Cloning of three new allergens from the dust mite Lepidoglyphus
RT destructor using phase surface display technology."
RL Eur. J. Biochem. 268:287-294(2001).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALLERGEN: Causes an allergic reaction in human.
CC -!- SIMILARITY: Belongs to the mite group 7 allergen family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ271058; CAB65963.1; -
CC DR EMBL; AJ271058; CAB65963.1; -
CC DR HAMAP; MF 00195; -; 1.
CC DR InterPro; IPR005289; GTP-binding_dom.
CC DR InterPro; IPR005073; GTP1_OBG.
CC DR InterPro; IPR002917; MNR_HSR1.
CC DR InterPro; IPR005225; Small_GTP.
CC DR Pfam; PF01926; MNR_HSR1; 1.
CC DR PRINTS; PR00326; GTP1_OBG.
CC DR TIGRFAMs; TIGR00650; MG442; 2.
CC DR TIGRFAMs; TIGR00231; small_GTP; 2.
CC KW GTP-binding; Repeat; Complete proteome.
CC FT NP_BIND 10 17 GTP 1 (POTENTIAL).
CC FT NP_BIND 57 61 GTP 1 (POTENTIAL).
CC FT NP_BIND 119 122 GTP 1 (POTENTIAL).
CC FT NP_BIND 181 188 GTP 2 (POTENTIAL).
CC FT NP_BIND 229 233 GTP 2 (POTENTIAL).
CC FT NP_BIND 294 297 GTP 2 (POTENTIAL).
CC FT CONFLICT 165 165 T -> A (IN REF. 1).
CC FT CONFLICT 311 311 A -> T (IN REF. 1).
CC FT CONFLICT 357 357 A -> S (IN REF. 1).
CC SQ SEQUENCE 216 AA; 23916 MW; 597LEAF05B4685E CRC64;
Query Match 23.5%; Score 250.5; DB 1; Length 216;
Best Local Similarity 29.2%; Pred. No. 2.7e-13;
Matches 63; Conservative 42; Mismatches 102; Indels 9; Gaps 5;
QY 2 MKELLTAA-VAFVAVSA--DPHYKITEEINKAIDDAIAAEQSETIDPMKVPDPAHX 57
Db 1 MOYLATAVIALAGLSAAAHKPAYD--DNMANQMDVQIVKSLTTTKKELDPFKIEQTKVP 58
QY 58 FERHVGIVDFKGLAMENIEARGLKQMKQGDANVKGEIGVKAHLIGVHDDIVSMYD 117
Db 59 IDKKIGLIHKSGATIKNAVITGLSHSRGDAKIDTDGGAFAATLKLG--DKNIRIKTD 116
QY 118 LAYKLG-DLPHTHVTSIDIOFVVALSLEISDGNITMTSFEVRQFANVNHIGLSILD 176
Db 117 LHLDLGKLIHPNLKFEHGIDGIDMKLKLDAEGKPSLDQFEIDPEQVELFHIGLGLD 176
QY 177 PIFGVLSDLVTAIFQDVTREKMTKVLAPAFKRELEK 212
Db 177 PLVDVIADSVFKVFNPOARKLVITDMLKPLIVBIEKK 212
RESULT 4
ID_ENGA_STRMU STANDARD; PRT; 436 AA.
AC QSDS30; Q9RHV5;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE GTP-binding protein engA.
GN ENGA OR SERA OR SMU.1920.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT8148 / Serotype C;
RX MEDLINE=21069259; PubMed=11155166;
RA Kawabata S., Terao Y., Hanada S.;
RT "Molecular cloning, sequence and characterization of a novel
RT streptococcal phosphoglycerate dehydrogenase gene.";
RL Oral Microbiol. Immunol. 15:58-62(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397185;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- FUNCTION: GTPase of unknown physiological role.
CC -1- SIMILARITY: Belongs to the era/trmE family of GTP-binding
CC proteins. EngA subfamily.
CC -1- CAUTION: Was originally (Ref.1) thought to be a D-3-
CC phosphoglycerate dehydrogenase.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB016077; BAA88823.1; -
CC DR EMBL; AB016077; BAA88823.1; -
CC DR HAMAP; MF 00195; -; 1.
CC DR InterPro; IPR005289; GTP-binding_dom.
CC DR InterPro; IPR005073; GTP1_OBG.
CC DR InterPro; IPR002917; MNR_HSR1.
CC DR InterPro; IPR005225; Small_GTP.
CC DR Pfam; PF01926; MNR_HSR1; 1.
CC DR PRINTS; PR00326; GTP1_OBG.
CC DR TIGRFAMs; TIGR00650; MG442; 2.
CC DR TIGRFAMs; TIGR00231; small_GTP; 2.
CC KW GTP-binding; Repeat; Complete proteome.
CC FT NP_BIND 10 17 GTP 1 (POTENTIAL).
CC FT NP_BIND 57 61 GTP 1 (POTENTIAL).
CC FT NP_BIND 119 122 GTP 1 (POTENTIAL).
CC FT NP_BIND 181 188 GTP 2 (POTENTIAL).
CC FT NP_BIND 229 233 GTP 2 (POTENTIAL).
CC FT NP_BIND 294 297 GTP 2 (POTENTIAL).
CC FT CONFLICT 165 165 T -> A (IN REF. 1).
CC FT CONFLICT 311 311 A -> T (IN REF. 1).
CC FT CONFLICT 357 357 A -> S (IN REF. 1).
CC SQ SEQUENCE 436 AA; 48585 MW; 24D8428A91C2A097 CRC64;
Query Match 9.6%; Score 103; DB 1; Length 436;
Best Local Similarity 23.0%; Pred. No. 0.48;
Matches 50; Conservative 36; Mismatches 77; Indels 54; Gaps 11;
QY 30 NKADDAIAAEQSETIDPMKVPDPAHXKFERHVGIVDFKGLAMENIEARGLKQMKRQGD 89
Db 21 NRIAGERISIVEDVEGVTDRDRIYTKAENLNFQSIIDTGG---IDDVDAPTFMEQIKHQAD 77
QY 90 AN-----VKGEIGV-----AHLIGVHDDIV-----SNEYDL-AYK 121
Db 78 IATMEADVIVFVVSNAKEITDADEVAKILRTHKPVILAVNKVDNPEWRSALIDFYALG 137
QY 122 LGDLHP--TTHVI--SDIQDFV-ALSLEISDEGNITWTSPEVRQF-----ANVV 166
Db 138 LGDPYVSSAHGIGTGVDLDAIVNLPTEAOBE-----SSDIKFSLIGRPNVKGKSLI 191
QY 167 NHIGLS--ILDPIFGVLSDLVTAIFQDVTREKMTKV.201
Db 192 NAILGEDRVIASPVAGTTRDAIDTTFDEBEGQFTMI 228
RESULT 5
ID_RPOC_CHLPN STANDARD; PRT; 1393 AA.
AC Q9Z999; Q9Z9J7; Q9K211;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR CFN0082 OR CF0693 OR CFN0082
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydiaophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RN Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,
RA White O., Hickey E.K., Peterson J., Otterback T., Berry K., Dodson R.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Salzborg S.L.,
RA Winn M., Nelson W., DeBoy R., Koonay J., McClarty G., Salzborg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TW-183;
RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: The enzyme consists of the sigma chain and the core
CC enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
CC beta' chain.
CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AE001593; AAD18235.1; -;
CC EMBL; AE002228; AAF38501.1; ALT_INIT.
CC EMBL; AF002545; BAA98292.1; -;
CC EMBL; AE017157; AAP98015.1; -;
CC PIR; B86501; B86501.
CC PIR; E72122; E72122.
CC HSSP; Q8KNU6; 1HQW.
CC TIGR; CP0693; -;
CC InterPro; IPR000722; RNA_pol_A.
CC InterPro; IPR007080; RNA_pol_Rpb1_1.
CC InterPro; IPR007066; RNA_pol_Rpb1_3.
CC InterPro; IPR007083; RNA_pol_Rpb1_4.
CC InterPro; IPR007081; RNA_pol_Rpb1_5.
CC InterPro; IPR006592; RNA_pol_N.
CC Pfam; PF04997; RNA_pol_Rpb1_1; 1.
CC Pfam; PF06623; RNA_pol_Rpb1_2; 1.
CC Pfam; PF04983; RNA_pol_Rpb1_3; 1.
CC Pfam; PF05000; RNA_pol_Rpb1_4; 1.
CC Pfam; PF04998; RNA_pol_Rpb1_5; 1.
CC SMART; SM00663; RPOA_N; 1.
CC Transferrase; DNA-directed RNA polymerase; Transcription;
CC Complete Proteome. 1031
FT CONFLICT 1031 AA; A -> G (IN REF. 1).
SQ SEQUENCE 1393 AA; 154900 MW; E0734EF236C6FDE8 CRC64;

Query Match
Best Local Similarity 21.2%; Pred. No. 16;
Matches 53; Conservative 35; Mismatches 63; Indels 99; Gaps 12;

QY 36 AIAAIEQSETIDP-----MKVPDHADKPERHVGIVD 66
Db 1106 AIISVEGQVDPGMLLAELPRGAIXKTDITGSLPRVAELVEARPEADADIAKIDGVVD 1165
QY 67 FKGEELAMRNI-----EARGLKQ-----MKRQGDANVKE---EGIVKAHLI--- 105
Db 1166 FKGIQNKRIILVVCDMTGMBEHLPLTKHLIVQKGDGVKQQLTDLGLVVPHEILETC 1225
QY 106 GYHD--DIVSMYDLAYKLGDLHPHTHVISDFQDFVAL-----SLEISDEGNITWTS 156
Db 1226 GYRELQKYLNEVQEVYRLQGV-----DINKHEIIVQMLQKVRITDPGDTILL- 1276
QY 157 FEVRQFANVYN-----HIGG-----LSILDPFGLVSDVITAFQDVTVR 195
Db 1277 -----FGEDVNKKEFVEENRTEEDGGKPAQVPLLGITKASLTGESFISAASQDT-- 1329
QY 196 KEMTKVLAPA 205
Db 1330 ---TRVLTD 1336

RESULT 6
CH60_XYLFA
ID CH60_XYLFA STANDARD; PRT; 547 AA.
AC Q9PF22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR XF0615.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5G;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Buco M.R.P., Camargo A.O., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Coutinho N.B., Colombo C., Costa P.P., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Mendck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).


```
RP SEQUENCE FROM N.A.
RC STRAIN-RIND 2210633 / Serotype O3:K6;
RA MEDLINE=22508454; PubMed=12620739;
RX Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RA "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
EL Lancet 361:743-749(2003).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF221845; AAF27528.1; -
DR EMBL; AF005083; BAC61114.1; -
DR HSSP; P06139; 1GRL.
DR HAMAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaprin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR InterPro; IPR008950; GROEL-ATPase.
DR Pfam; PF00118; Cpn60_TCF1; 1.
DR PRINTS; PRO0298; CHAPERONIN60.
DR PRINTS; PRO0304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR Chaperone; ATP-binding; Complete proteome.
KW CONFLICT 545 545 G -> GG (IN REF. 1).
FT CONFIDENCE 547 AA; 57572 MW; 1220289D6F74CES CRC64;
SQ
Query Match 8.5%; Score 90.5; DB 1; Length 547;
Best Local Similarity 20.9%; Pred. No. 6.4;
Matches 49; Conservative 40; Mismatches 82; Indels 63; Gaps 9;
QY 13 VAVSADPHYKITEINKAIDDAI-AAIEQSEIDPMKVPDPHADKFERHVGIVDFKGL 71
DB 107 VAAAGNPM-----DLKRGIDKAAVAVEQ---LKELSVENCTKRAIQVGTISANSDA 156
QY 72 AMRNTAEAGLQKMKRQGDANVKGEEGVKAHLIGVHDDIVSMEDYLAY-----120
DB 157 SVGNIIAEMERVGEDGVIIV-EG-----QALQDELVDVGMGFDGRGYLSPYFINNQE 210
QY 121 -----KLGLDHTTHVDSIDIQDFVVALSLEISDGNITMTSFEVRQF 162
DB 211 GSVELENFILLVDKIKSIRELLPTLEAVAKASRPILLITAEVDVEGALATL-----262
QY 163 ANVNVHIGGL---SILDPFGVLSDLVLTATFQTVRKMTKVLAFAFKRELEX 212
DB 263 --VNNMGIVKVAAPKPGFG---DRKAMLDQIALITGTVISEIGLELEX 311
RESULT 9
Y373 HUMAN
ID Y373 HUMAN STANDARD; PRT; 1539 AA.
AC O15078.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=93066366; PubMed=1279700;
RA Peattie D.A., Harding M.W., Fleming M.A., Decenzo M.T.,
RA Lipkpe J.A., Livingston D.J., Benasutti M.;
RT "Expression and characterization of human FKBP52, an immunophilin
RT that associates with the 90-kDa heat shock protein and is a component
RT of steroid receptor complexes.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10974-10978(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, Lymph, and Uterus;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB002371; BAA20828.1; -
DR Hypothetical protein; Coiled coil.
FT DOMAIN 18 1514
SQ SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;
Query Match 8.4%; Score 89.5; DB 1; Length 1539;
Best Local Similarity 18.8%; Pred. No. 26;
Matches 29; Conservative 43; Mismatches 67; Indels 15; Gaps 3;
QY 3 KELLTAAVAFVAVSADPHYDKITEINKAIDDA-----IAAIEQSE---TIDPMKVPDHA 55
DB 160 KFAETTKINLDQAKVEQMLRDLADSVKAVSDAQRILKELKNEMLKVEVSKLRIS 219
QY 56 DKFERHVGIVDFKGLAMENIEARGLKQMKRQGDANVKGEEGVKAHLIGVHDDIVSME 115
DB 220 DIARRQVELNNAQQSRDKESVLSMQLLDYQAQSDKS-----LIAKLQHNVSLQ 271
QY 116 YDLAYKGLDHTTHVDSIDIQDFVVALSLEISDE 149
DB 272 LSEATALGKLEISITKIQKMEAYNLRLQKLEK 305
RESULT 10
FKB4 HUMAN
ID FKB4 HUMAN STANDARD; PRT; 458 AA.
AC Q02790; Q9UCU1; Q9UCV7;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE FK506-binding protein 4 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans
DE isomerase) (PPIase) (Rotamase) (p59 protein) (HSP binding
DE immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein)
DE (FKBP59).
GN FKBP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=93066366; PubMed=1279700;
RA Peattie D.A., Harding M.W., Fleming M.A., Decenzo M.T.,
RA Lipkpe J.A., Livingston D.J., Benasutti M.;
RT "Expression and characterization of human FKBP52, an immunophilin
RT that associates with the 90-kDa heat shock protein and is a component
RT of steroid receptor complexes.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10974-10978(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, Lymph, and Uterus;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

DR	ENML; BC002887; AAH02887.1; -.
DR	ENML; BC007924; AAH07924.1; -.
DR	PIR; A46372; A46372.
DR	PDB; INIA; 30-DEC-02.
DR	Genew; HGNC:3720; FKBP4.
DR	MIM; 600611; -.
DR	GO; GO:0005737; C:cytosolasm; TAS.
DR	GO; GO:0005528; F:FK506 binding; TAS.
DR	GO; GO:0008077; F:Hsp70/Hsp90 organizing protein activity; TAS.
DR	GO; GO:0006457; P:protein folding; TAS.
DR	InterPro; IPR001179; FKBP_PPIase.
DR	InterPro; IPR008941; TPR-like.
DR	InterPro; IPR001440; TPR.
DR	pfam; PF00254; FKBP; 2.
DR	pfam; PF00515; TPR; 3.
DR	SMART; SM00028; TPR; 3.
DR	PROSITE; PS00453; FKBP_PPIASE 1; 1.
DR	PROSITE; PS00454; FKBP_PPIASE 2; 2.
DR	PROSITE; PS00059; FKBP_PPIASE 3; 2.
KW	Isomerase; Rotamase; Repeat; TPR repeat; Nuclear protein;
KW	Phosphorylation; 3D-structure.
FT	INIT_MET 0
FT	DOMAIN 49 137 PPIASE, FKBP-TYPE 1.
FT	DOMAIN 166 252 PPIASE, FKBP-TYPE 2.
FT	REPEAT 269 302 TPR 1.
FT	REPEAT 318 351 TPR 2.
FT	REPEAT 353 385 TPR 3.
FT	MOD_RES 142 142 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT	CONFLICT 145 146 ED -> AR [IN REF. 2; AAH02887].
FT	SEQUENCE 458 AA; 51673 MW; 916B3B945CS1634E CRC64;
QY	Query Match 8.3%; Score 88.5; DB 1; Length 458;
Db	Best Local Similarity 24.6%; Pred. No. 7.5;
Db	Matches 30; Conservative 25; Mismatches 50; Indels 17; Gaps 3
QY	18 DLIHYDKITEINKAIDDAIAAEQSFI-----DPMKVDPHAD-KPERHV 62
Db	74 DKFSFDLGKGEIVIKAWDJAIATMKVGVECHITCKPEYAYGSAGSPPKIPPNAFLVE--V 131
QY	63 GLVDFKGLAMNTEARGLKQMKGQDANVKGEGIVKAHLIGVHDDIVSMEYDLAVKL 122
Db	132 ELFRFGEDLTREDGGIIIRIQRGRGYAKPNEGAVEVALEGGYKYKLFQRELRFEI 191
QY	123 GD 124
Db	192 GE 193
RESULT 11	
Y373_BOVIN	
ID Y373 BOVIN STANDARD; PRT; 1453 AA.	
AC Q9TU23;	
DT 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DE 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Hypothetical protein KIAA0373.	
GN KIAA0373.	
OS Bos taurus (Bovine).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC Bovidae; Bovinae; Bos.	
OX NCBI_TaxID=9913;	
[1]	
RN SEQUENCE FROM N.A.	
RP TISSUE=Trachea;	
RA Jovov B., Ripoll P.J., Benos D.J.;	
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration-	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 CC DR: EMBL; AF176816; AAF00990.1; -
 CC KW Hypothetical protein; Coiled coil.
 CC FT DOMAIN 37 1426 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF139 CRC64;

Query Match 8.3%; Score 88.5; DB 1; Length 1453;
 Best Local Similarity 17.5%; Pred. No. 29;
 Matches 27; Conservative 43; Mismatches 69; Indels 15; Gaps 3;

QY 3 KELLIAAVFAVAVSADPHYDKITEINKAIDDA-----IAAIEOSE---TIDPMKVPDHA 55
 DB 179 KPAELTRINLEAQAQVEQMLRDELADSVSKTVSDADROHILELSEKSELMELKVEVSKLEIS 238
 QY 56 DKFERHVGIVDFKGLAMRNIEARGLKQKRGDANVKGEIGVKAHLIGVHDDIVSME 115
 DB 239 DIAKQVEILNAQQQSREKEVESLRTQLLDYQAQSDKA-----LIAKLHQHVVSQ 290
 QY 116 YDLAVKGLDLPHTTHVTSIDIQDFVVALSLEISDE 149
 DB 291 ASEAALGKVESVASKLQKVEAHTLRLEQKLDEX 324

RESULT 12

SPED_BUCAP STANDARD; PRT; 268 AA.
 AC Q8K9T6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdoMetDC)
 DE (SamDC) [Contains: S-adenosylmethionine decarboxylase beta chain; S-adenosylmethionine decarboxylase alpha chain].
 GN SPED OR BUSG202.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=98794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22084549; PubMed=12089438;
 RA Tamas I., Klaassen L., Canbaeck B., Naeslund A.K., Eriksson A.-S., Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";
 RL Science 296:2376-2379 (2002).
 CC -!- FUNCTION: Decarboxylation of S-adenosylmethionine provides the aminopropyl moiety required for spermidine biosynthesis from putrescine (By similarity).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine = (S)-deoxy-S-adenosyl (3-aminopropyl) methylsulfonium salt + CO(2).
 CC -!- COFACTOR: Pyruvoyl group (By similarity).
 CC -!- SIMILARITY: Belongs to the prokaryotic AdoMetDC family. Subfamily 2.

 This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL; AE014096; AAM67766.1; -
 CC DR HAMAP; MF_00465; -; 1.
 CC DR InterPro; IPR003826; SAMDC.
 CC DR Pfam; PF02675; AdoMetDC; 1.
 CC DR PIRSF; PIRSF001356; SAM decarboxylase; 1.
 CC KW Spermidine biosynthesis; Lyase; Decarboxylase; Zymogen; Pyruvate;
 CC Complete proteome.
 FT CHAIN 1 112 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA CHAIN (BY SIMILARITY).

FT CHAIN 113 268 S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA CHAIN (BY SIMILARITY).
 FT SITE 112 113 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).
 FT MOD_RES 113 CONVERTED TO A PYRUVOYL GROUP (BY SIMILARITY).
 FT SQ SEQUENCE 268 AA; 31252 MW; 69A5F94715D2CC57 CRC64;

Query Match 8.2%; Score 88; DB 1; Length 268;
 Best Local Similarity 21.7%; Pred. No. 44;
 Matches 47; Conservative 38; Mismatches 64; Indels 68; Gaps 11;

QY 13 VAVSADPHYDK--ITEINKAIDDAIAIEQSEITIDPMKVPDHAKEF-----ERH- 61
 DB 80 ILVCEEPINLEKVNKSNKISSVLA-----HLDKSHCVHTVPESH 125
 QY 62 -VGIVDFKGLAMRNIEARGLKQKRGDANVKGEIGVKA-----HLLGVHDDIVSME 115
 DB 126 QSGICTFRADIEVSTC-----GIISPLNALNYLIHQLESIVTIE 165
 QY 116 YDLAVKGLDLPHTTHV---ISDIQDFV---VALSLEISD---EGNITMTSFVEVRQFAN 164
 DB 166 YVVRGETRIDHIGKHFIDHKINSIQNFMSHDKSMYEMFDVNIYQENIFHTRMLKEF-N 224
 QY 165 VVNHHGGISILDPFGVLSVLTATFQDVTVRKEMTKV 201
 DB 225 LKNYLFNINVKD-----LSKKEHSYIIDLLWKEMREI 256

RESULT 13

ENGA_STRPN STANDARD; PRT; 436 AA.
 AC Q97PC9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE GTP-binding protein enga.
 GN ENGA OR SERA OR SP1709 OR SPRI553.
 OS Streptococcus pneumoniae, and
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313, 171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-334 / TIGR4;
 RC MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D., Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., E.K., McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Doughterty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
 RL Science 293:498-506 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-255 / R6;
 RC MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S., DeHoff B.S., Bateman S.T., Fritz L., Fu D.-J., Fuller W., Geringer C., Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E., LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P., McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicot T.I., Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P., Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G., Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L., Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717 (2001).
 CC -!- FUNCTION: GTPase of unknown physiological role.

Q9AJ82;
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
 GN (Protein PS 120) (Fragment)
 OS Rickettsia mongolotimonae.
 OC Rickettsia: Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 RX NCBI_TaxID=45261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sekoyova Z., Roux V., Raoult D.;
 RT 'Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
 RL 'gene D' coding for an intracytoplasmic protein';
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF151725; AAK30683.1; -;
 KW Antigen.
 FT NON_TER 1 1011
 FT NON_TER 1011 1011
 FT NON_TER 1011 1011
 SQ SEQUENCE 1011 AA; 110607 MW; 0169A06981BDS08 CRC64;
 Query Match 8.2%; Score 87.5; DB 1; Length 1011;
 Best Local Similarity 22.4%; Pred. NO. 23;
 Matches 55; Conservative 35; Mismatches 98; Indels 57; Gaps 10;
 QY 5 LLIAAVAFVAVSADPIHYDK--ITEEI--NKAID---DATAIEQSETIDPMKVPDADK 57
 DB 402 LLNTATLGGSMQDLNLYNAGLTKEIDSNKQIDLIKEAATAILNNE-----KSDI 452
 QY 58 FERHVGIVDFKGEIAMRNIIEARGLKQMKQDQANVKGEGIV-----KAHL 104
 DB 453 AEKQANII-----ALAEVTNNKLP-----DAKVGAVNAVLETIKNDQTPNLEKML 503
 QY 105 IGVEHDDIVSME-----YDLAYKLGDLHPTTHVISDIQDFVVALSLEISDEG 150
 DB 504 EATVAIILNSENLEPKQKQOMLEKXAVDGLSLKDDASRAVTIDGKDVVIXTNLSTEDKG 563
 QY 151 NITWT---SPFVROFANVNVHIGLSILDPITFGVLSVLTAIFQDVTVRKMTKVLAPFK 207
 DB 564 TMLTAGVDGKVVNSLSNAEKSKSGSVLKK--GVEAQVLSPAQQLIQHLDKITAEOIK 621
 QY 208 RELEK 212
 DB 622 KDTIK 626
 RESULT 16
 ID ENGA STRP3 STANDARD; PRT; 436 AA.
 AC Q8K8J8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE GTP-binding protein engA.
 GN ENGA OR PGDA OR SPYM3_0249 OR SP31610.
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 RX NCBI_TaxID=198466;
 RN [1]
 RP SEQUENCE FROM N.A.

STRAIN=MGAS315 / Serotype M3;
 MEDLINE=22133808; PubMed=12122206;
 RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
 RA Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
 RT phage-encoded toxins, the high-virulence phenotype, and clone
 RT emergence";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=SSI-1 / Serotype M3;
 MEDLINE=22683278; PubMed=12799345;
 RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
 RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
 RA Hayashi H., Hattori M., Hamada S.;
 RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
 RT large-scale genomic rearrangement in invasive strains and new insights
 RT into phage evolution";
 RL Genome Res. 13:1042-1055(2003).
 CC -!- FUNCTION: GTPase of unknown physiological role.
 CC -!- SIMILARITY: Belongs to the era/trme family of GTP-binding
 CC proteins. EngA subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB014141; AM78856.1; -;
 DR EMBL; AP005146; BAC64705.1; -;
 DR HAMAP; MF_00195; -; 1.
 DR InterPro; IPR005289; GTP-binding_dom.
 DR InterPro; IPR006073; GTP_OBG.
 DR InterPro; IPR002917; MMR_HSR1.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF01926; MMR_HSR1; 1.
 DR PRINTS; PR00326; GTP_OBG.
 DR TIGRFAMs; TIGR00650; MG442; 2.
 DR TIGRFAMs; TIGR00231; small_GTP; 2.
 KW GTP-binding; Repeat; Complete proteome.
 FT NP_BIND 10 17 GTP 1 (POTENTIAL).
 FT NP_BIND 57 61 GTP 1 (POTENTIAL).
 FT NP_BIND 119 122 GTP 1 (POTENTIAL).
 FT NP_BIND 181 188 GTP 2 (POTENTIAL).
 FT NP_BIND 229 233 GTP 2 (POTENTIAL).
 FT NP_BIND 294 297 GTP 2 (POTENTIAL).
 SQ SEQUENCE 436 AA; 48771 MW; F2DF6CD5B674323A CRC64;
 Query Match 8.1%; Score 87; DB 1; Length 436;
 Best Local Similarity 22.2%; Pred. NO. 9.4;
 Matches 47; Conservative 41; Mismatches 80; Indels 44; Gaps 11;
 QY 30 NKAIDDAIAAQSETIDPMKVPDADKFERHVGIVDFKGEIAMRNIIEARGLKQMKRQG- 88
 DB 21 NRIAGERISIVDEVEGTRDRIYATGEWLNQFSLDTGG---IDDDVAPMEQIKQQAQ 77
 QY 89 -----DAN-----VKGEEGIVKAH-----LLIGVHD-DIVSMEYDL-----AYK 121
 DB 78 IAMEEADVIVFVSGKEGVTDADEVSKILYRTNTPVILAVNKVDNPMRNDIVDFYSLG 137
 QY 122 LGDLHPTTHV---ISDIQDFVVA--LSLEISDEGNITMTSEFEVQFANV-----VNHIGG 171
 DB 138 LGDPYFVSSVSGIGTGDVDAIVENLPVEBAEE--NDDIRFSLIGRPNVNGKSSLINAILG 196
 QY 172 LS--ILDPFGVLSVLTAIFAIFQDVTVRKMTKV 201
 DB 197 EDRVIASPVAGTTRDAIDTHTFDADGQEFTMI 228

RESULT 17
 ID ENGA STRPY STANDARD; PRT; 436 AA.
 AC Q9ALD0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE GTP-binding protein, enga.
 GN ENGA OR P6DA OR SPY0341 OR SPY18_0334.
 OS Streptococcus pyogenes, and
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314, 186103;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savić D.J., Savić G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar P.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
 RN [2]
 SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=1917108;
 RA Smoot J.C., Barbhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.B., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 group A *Streptococcus* strains associated with acute rheumatic fever
 outbreaks";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
 CC -1- SIMILARITY: GTPase of unknown physiological role.
 CC -1- SIMILARITY: Belongs to the era/time family of GTP-binding
 proteins. ENGA subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AE006498; AKK33393.1; -;
 DR EMBL; AE009978; AAL97088.1; -;
 DR HAWAP; MF_00195; -; 1.
 DR InterPro; IPR005289; GTP-binding_dom.
 DR InterPro; IPR006073; GTP1_OBG.
 DR InterPro; IPR002917; MMR_HSR1.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF01926; MMR_HSR1; 1.
 DR PRINTS; PR00326; GTP1_OBG.
 DR TIGRFAMs; TIGR00650; MG442; 2.
 DR TIGRFAMs; TIGR00231; small_GTP; 2.
 KW GTP-binding; Repeat; Complete proteome.
 FT NP_BIND 10 17 GTP 1 (POTENTIAL).
 FT NP_BIND 57 61 GTP 1 (POTENTIAL).
 FT NP_BIND 119 122 GTP 1 (POTENTIAL).
 FT NP_BIND 181 188 GTP 2 (POTENTIAL).
 FT NP_BIND 229 233 GTP 2 (POTENTIAL).
 FT NP_BIND 294 297 GTP 2 (POTENTIAL).
 SQ SEQUENCE 436 AA; 48901 MW; E9DED7D51675323A CRC64;

Query Match 8.18; Score 87; DB 1; Length 436;

Best Local Similarity 22.28; Pred No. 9.4; Mismatches 44; Gaps 11;

QY 30 NKAIDDAIAAEQSETIDPMKVPDADKFEHVGIVDFKGLAMRNIEARGLKQKRG- 88
 DB 21 NRIAGERISIVEDVEGVRTRDRIYATGEWLNQFSLDTGG---IDDVDPMEQIKHQAG 77
 QY 89 ----DAN-----VKGEIGIVKAH-----LLIGVHD-DIVSMEYDL-----AYK 121
 DB 78 IAMEEADVIVFVSGKGVTDADDEVSKILYRTNTFTVILAVNKVDNPEMRNDIVDFYSLG 137
 QY 122 LGDLHPPTHV-----ISDIQDFVVA-LSLEISDEGNITWTSFEVRFQFANY-----VNHGG 171
 DB 138 LGDPYVSSVHGIGTGDVLDIAENLPVEAAE-NDDIIRFSLIGRPNVGKSLINAILG 196
 QY 172 LS--ILDIPIFGLVSLTAIFQDITVRKEMTKV 201
 DB 197 EDRVIASPVAGTTRDAIDTHTDADGQEFMTI 228
 RESULT 18
 ID FKBP4 MOUSE STANDARD; PRT; 457 AA.
 AC P30416;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE FK506-binding protein 4 (EC 5.2.1.8) (Peptidyl-prolyl cis-trans
 DE isomerase) (PPIase) (Rotamase) (p59 protein) (HSP binding
 DE immunophilin) (HBI) (FKBP52 protein) (52 kDa FK506 binding protein)
 DE (FKBP59).
 GN FKBP4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=94040772; PubMed=7693550;
 RA Schmitt J., Stummenberg H.G.;
 RT "Cloning and expression of a mouse cDNA encoding p59, an immunophilin
 RT that associates with the glucocorticoid receptor";
 RL Gene 132:267-271 (1993).
 RN [2]
 SEQUENCE FROM N.A.
 RP TISSUE=Lung tumor;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.N., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 SEQUENCE OF 5-457 FROM N.A., PARTIAL SEQUENCE, FUNCTION, AND
 RP SUBCELLULAR LOCATION.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=93342084; PubMed=8341706;
 RA Alnemri E.S., Fernandes-Alnemri T., Nelki D.S., Dudley K.,
 RA Dubois G.C., Litwack G.;
 RT "Overexpression, characterization, and purification of a recombinant
 RT mouse immunophilin FKBP-52 and identification of an associated

phosphoprotein.";
Proc. Natl. Acad. Sci. U.S.A. 90:6839-6843(1993).
-!- FUNCTION: Component of unactivated mammalian steroid receptor
complexes that sediment at 8-10 S. May have a rotamase activity.
May play a role in the intracellular trafficking of hetero-
oligomeric forms of steroid hormone receptors.
-!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
(omega=0).
-!- SUBUNIT: Associates with HSP90 and HSP70 in unactivated steroid
hormone receptor complexes, also interacts with peroxisomal
phytanoyl-CoA alpha-hydroxylase (PHYH). Interacts with NR3C1 and
dyx1n (By similarity).
-!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
-!- PTM: Phosphorylation by CK2 results in loss of HSP90 binding
activity (By similarity).
-!- SIMILARITY: Belongs to the FKBP-type PPIase family.
-!- SIMILARITY: Contains 2 FKBP-type PPIase domains.
-!- SIMILARITY: Contains 3 TPR repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X70887; CAAS0231.1; -.
EMBL; BC003447; AAH03447.1; -.
EMBL; X17069; CAC39452.1; -.
EMBL; X17068; CAA34914.1; ALT_SEQ.
PIR; JN0873; JN0873.
HSP; F27124; IROT.
MGD; MGI:95543; Fkbp4.
InterPro; IPR001179; FKBP_PPIase.
InterPro; IPR008941; TPR-like.
InterPro; IPR001440; TPR.
Pfam; PF00254; FKBP; 2.
Pfam; PF00515; TPR; 3.
SMART; SM00028; TPR; 3.
PROSITE; PS00453; FKBP_PPIASE_1; 1.
PROSITE; PS00454; FKBP_PPIASE_2; 2.
PROSITE; PS00509; FKBP_PPIASE_3; 2.
IsoMase; Rotamase; Repeat; TPR repeat; Nuclear protein;
Phosphorylation.
INIT MET 0 BY SIMILARITY.
FT DOMAIN 49 137 PPIASE, FKBP-TYPE 1.
FT DOMAIN 166 252 PPIASE, FKBP-TYPE 2.
FT REPEAT 269 302 TPR 1.
FT REPEAT 318 351 TPR 2.
FT REPEAT 352 385 TPR 3.
FT MOD_RES 142 142 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT MOD_RES 5 6 MK -> HE (IN REF. 3).
FT CONFLICT 155 162 TRGGYAR -> LGVKMQG (IN REF. 3).
FT CONFLICT 202 239 GLEBAIQRMKGESIVLKPSYAFSGVGRKFOIPPH ->
AWRPFSARWKEPSCTSNLMLLVGRRGSRSHRT
(IN REF. 3).
FT CONFLICT 264 271 EKLQSNL -> RQWSRATY (IN REF. 3).
FT CONFLICT 314 314 H -> R (IN REF. 2).
FT SEQUENCE 457 AA; 51441 MW; 0D087C543541C1C0 CRC64;

Query Match 8.1%; Score 86.5; DB 1; Length 457;
Best Local Similarity 23.0%; Pred. No. 11;
Matches 28; Conservative 27; Mismatches 50; Indels 17; Gaps 3;

QY 18 DPVHYDKITEINKAIDAAIAEQSETI-----DPMKVPDHAD-KFERHV 62
DB 74 DKFSDLKGGEVIRKWDIAVATMKVGEVCHTCKPEYAYGAGSPFKIPPNATVFE--V 131

QY 63 GIVDFKGLAMNTEIARGLKMKQGGDANVKGEGIVKHAHLLIGVDDIVSMYVDLAYKL 122
DB 132 ELFFPKGDLTEEDGGIIRIRTRGEGYARPDNGAMVEALEGYHKDRUFDQRELCEV 191

QY 123 GD 124
DB 192 GE 193

RESULT 19
CH61_VIBCH STANDARD; PRT; 544 AA.
ID CH61_VIBCH
AC Q9KNR7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin 1 (protein Cpn60 1) (GroEL protein 1).
GN GROEL1 OR GROEL1 OR VC2664.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Bimolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
proper assembly of unfolded polypeptides generated under stress
conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE004332; AAF95805.1; -.
PIR; B82048; B82048.
HSP; P06139; 1GRL.
TIGR; VC2664; -.
HAMAP; MF_00600; -; 1.
InterPro; IPR001844; Chaperin Cpn60.
InterPro; IPR002423; Cpn60/TCP-1.
InterPro; IPR008950; GroEL-ATPase.
Pfam; PF00118; Cpn60_TCP1; 1.
PRINTS; PR00298; CHAPERONING60.
PRINTS; PR00304; TCOMPLEXTCP1.
PROSITE; PS00296; CHAPERONINS_CPN60; 1.
Chaperone; ATP-binding; Complete proteome.
SEQUENCE 544 AA; 57153 MW; 927868AE794AE079 CRC64;

Query Match 8.1%; Score 86.5; DB 1; Length 544;
Best Local Similarity 24.4%; Pred. No. 13;
Matches 41; Conservative 21; Mismatches 69; Indels 37; Gaps 5;

QY 6 LTAANFAVAVSR-----DPIHYDKITEINKAIDAAIAEQSETIDPMKV 51
DB 372 LAGGVAIVKVGAAATEVEMKEKDRVEDDAHATRAAEEGVAGGVALLIRAAKLSLV- 430

QY 52 PDHAKDFERHVGIVDFKGLAMNTEIARGLKMKQGGD-----NVKGEIGIVKHAHLLI 105


```
Db 431 ---GDNEQNGVI-----RVALRAMEAPLRQIVKNAGDEBSVANNVRAGEGNYGNAAT 482
Qy 106 GVHDDIVSMYDLAYKGLDHPHTTHVISDIQDFVVALSLSDSGNIT 153
Db 483 GVGDMTEM-----GILDPKTVRSALQFAASVAGLMTTTEAMIT 522

RESULT 20
CH60_BUCTT STANDARD; PRT; 553 AA.
AC Q8KIXI;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL.
OS Buchnera aphidicola (subsp. Tuberolachnus salignus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98804;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=2077640; PubMed=12082135;
RA Fares M.A., Barrio E., Sabater-Munoz B., Moya A.;
RT "The evolution of the heat-shock protein GroEL from Buchnera, the
RT primary endosymbiont of aphids, is governed by positive selection.";
RL Mol. Biol. Evol. 19;1162-1170(2002).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ439086; CAD27798.1; -.
DR HAMAP; MF 00600; -.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60_TCF1.1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PROSITE; PS00304; TCOMPLEXTCPL.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 553 AA; 58010 MW; 814F300472676701 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 553;
Best Local Similarity 18.6%; Pred. No. 16;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

Qy 13 VASADPIHYDKITEINKAIDDA-IAAIEQSETIDPMKVPDHADKFERHVGIVDFKGL 71
Db 107 VAAGMFWP-----DLKRGIDKAVIAVEE---LRKLSVPCSDSKAIAQVGTISANDL 156

Qy 72 AMRNIARGLKQMKRQGDANVKGEIGVKAHLIGVHDD---IVSMYDLAY-----120
Db 157 TVGKLIAEAMGKV-----GKEGIVTVEEGTGLQDELVDVVEGQFQFGYLSFYFINK 207

Qy 121 -----KLGDLPHTTHVLSIDIQDFVVALSLSDSGNITMTSFEV 159
Db 208 PESGALENPFIILADKKISNRELLPFILEAVAKAGKSLIIIAEDVEGEALATL-----262

Qy 160 RQFANVNVHIGGL-----SILEPIFGVLSVLTAFQDTVRKENTKVLAPAFKEELEK 212
Db 262 -----SILEPIFGVLSVLTAFQDTVRKENTKVLAPAFKEELEK 212

Db 263 -----IVNTRMGIVKVAAPKAPFG---DRRKAMLDIATLTGGTVISEEIGLELEK 311

RESULT 21
YB56_METJA STANDARD; PRT; 903 AA.
AC Q58556;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cell division cycle protein 48 homolog MJ1156.
GN MJ1156.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273;1058-1073(1996).
CC -1- SIMILARITY: Belongs to the AAA ATPase family. CDC48 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67557; AAS99153.1; -.
DR FIR; C64444; C64444.
DR TIGR; MJ1156; -.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003959; AAA ATPase_cent.
DR InterPro; IPR005938; AAA CDC48.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR009010; Asp_decarb fold.
DR InterPro; IPR003338; ATPaseVAT_N.
DR InterPro; IPR004201; Cdc48_2.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02933; cdc48_2; 1.
DR Pfam; PF02359; cdc48_N; 1.
DR SMART; SM00382; AAA_2.
DR TIGRfams; TIGR01243; CDC48; 1.
DR PROSITE; PS00574; AAA; 2.
KW Hypothetical protein; ATP-binding; Repeat; Multigene family;
KW Complete proteome.
FT NP_BIND 220 227 ATP (POTENTIAL).
FT NP_BIND 493 500 ATP (POTENTIAL).
SQ SEQUENCE 903 AA; 100403 MW; 99F2ABEE2D544DC2 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 903;
Best Local Similarity 25.0%; Pred. No. 29;
Matches 53; Conservative 30; Mismatches 90; Indels 39; Gaps 11;

Qy 23 DKITEINKAIDD-----AIAAIEQSETIDPMKV-PHADKFERHVGIVDFKGLAM 73
Db 571 DKVNNQLTELDGMEEPKDVVVVIAATNPDIIDPALLRPGRLDRVIL-VPVDEKARLDI 629

Qy 74 RNIEARGLKQMKRQGDANV---KGEIGVKAHL-----LIGVHDDIVSMYDLAYK 121
Db 121 -----LIGVHDDIVSMYDLAYK 121
```

Db 630 FKIHTR----SMNLADVNLELAKKTEGYTGADIEALCREAAMLAVERSI-GKPDWIEVK 695
 QY 122 LGDLHPTTHVTSIDQDFVVALSLSEIDEGNIT--MTSPFVROQFANVNVHIGGLSILDPF 179
 Db 696 LREL---INYLQISGTFRAAAVELNSVTKATKERSAAGEFSELKNAIG-----KII 736
 QY 180 GVLSVLTALFQDVTVRKEMTKVLAPAFKRELE 211
 Db 737 SVLSPAKEKI--EAVEKEIDKFLVEINKEELK 766

RESULT 22
 RPOC_CHLMU
 ID RPOC_CHLMU STANDARD; PRT; 1396 AA.
 AC Q9PK79;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
 beta' chain) (RNA polymerase beta' subunit).
 GN RPOC OR TC0588.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MoPn / Ni99;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J.K., Khouri H., Craven B., Bowman R.,
 RA Gwinn M., Nelson C.M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 of DNA into RNA using the four ribonucleoside triphosphates as
 substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 CC -!- SUBUNIT: The enzyme consists of the sigma chain and the core
 enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
 beta' chain.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF002327; XAF39420.1; --
 DR PIR; F81686; F81686.
 DR HSSP; Q9KWU6; 1HQW.
 DR TIGR; TC0588; --
 DR InterPro; IPR007022; RNA_pol_A.
 DR InterPro; IPR007080; RNA_pol_Rpb1_1.
 DR InterPro; IPR007066; RNA_pol_Rpb1_3.
 DR InterPro; IPR007083; RNA_pol_Rpb1_4.
 DR InterPro; IPR007081; RNA_pol_Rpb1_5.
 DR InterPro; IPR006592; RNA_pol_N.
 DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
 DR Pfam; PF06623; RNA_pol_Rpb1_2; 1.
 DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
 DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
 DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
 DR SMART; SM00663; RPOLA_N; 1.
 DR Transferrase; DNA-directed RNA polymerase; Transcription;
 KW Complete proteome.
 SQ SEQUENCE 1396 AA; 154897 MW; CC090B9A18986977 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 1396;
 Best Local Similarity 23.6%; Pred. No. 48;
 Matches 49; Conservative 34; Mismatches 66; Indels 59; Gaps 12;
 QY 43 STIDPMKVPDHDKFERHVGIVDFKGLAMENI-----EARGLKO-----MKRQ 87
 Db 1145 AELVEARPEDAAD-IAKIDGVVDFKQKKNRIILVWDEVTGMEEEHLISLTKHLIVQR 1203
 QY 88 GDANVKGE---EGIVKAHLI---GVHD--DIVSMYDLAYKLGDLHPTTHVISDIOFV 139
 Db 1204 GDSVIKGOQLTDGLVVPHEILLEICGRELQKVLNNEVQEVYRLQGV-----DINDKH 1255
 QY 140 VAL-----SLEISDEGNITM---TSFEVROFANV---VNHIGG-----LSILDP 177
 Db 1256 IEIIVRQMLQKVRITDPGDTLLFGEDVDKKEFYBENRRTTEDGGKPAQAVPVLGITKA 1315
 QY 178 IFGVLSVLTALFQDVTVRKEMTKVLAPA 205
 Db 1316 SLGTESFISAAFSQDT-----TRVLDTA 1338

RESULT 23
 RPOC_CHLTR
 ID RPOC_CHLTR STANDARD; PRT; 1396 AA.
 AC Q843I6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
 beta' chain) (RNA polymerase beta' subunit).
 GN RPOC OR CT314.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/Cx;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
 of DNA into RNA using the four ribonucleoside triphosphates as
 substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 CC -!- SUBUNIT: The enzyme consists of the sigma chain and the core
 enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
 beta' chain.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF003304; AAC67907.1; --
 DR PIR; G71529; G71529.
 DR HSSP; Q9KWU6; 1HQW.
 DR PHCI-2DPAGE; O84316; --
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR007080; RNA_pol_Rpb1_1.
 DR InterPro; IPR007066; RNA_pol_Rpb1_3.
 DR InterPro; IPR007083; RNA_pol_Rpb1_4.
 DR InterPro; IPR007081; RNA_pol_Rpb1_5.
 DR InterPro; IPR006592; RNA_pol_N.

DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR SMART; SM00663; RPOA_N; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Complete procuome.
SQ SEQUENCE 1396 AA; 154904 MW; B24BF9841D284065 CRC64;

Query Match 8.0%; Score 85.5; DB 1; Length 1396;
Best Local Similarity 24.0%; Pred. No. 48;
Matches 50; Conservative 33; Mismatches 66; Indels 59; Gaps 12;

QY 43 SETIDPMKVPDPHADKFERHVGIVDPKGLAMRNIT-----BARGLKQ-----MKRQ 87
DB 1145 AELVEAKPEADAD-IAKIDGWDFKGIQNKRIILVVRDEITGMEEEHLISLTKHLIVQR 1203

QY 88 GDANVKEG--EGIVKAHLII--GVHD--DIVSMYDYLAKGLDHLPTTHVISDIQDFV 139
DB 1204 GDSVIKQQQLTDGLVPHLEICGVRELQKYLNVQEVYRQGV-----DINDKH 1255

QY 140 VAL-----SLEISDSGNITM--TSFVRQFANV---VNHIGG-----LSILDP 177
DB 1256 VEITVRMLQKVRITDPDITLLFGEDVDKKEFYENRTEEDGKPAQVVPVLLGITKA 1315

QY 178 IFGVLSVLTALFQDVTVRKEMTKVLAPA 205
DB 1316 SLGTESFISAASFDIT-----TRVLTDA 1338

RESULT 24
CH60_BUCTS
ID CH60_BUCTS STANDARD; PRT; 550 AA.
AC Q8KX2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL.
OS Buchnera aphidicola (subsp. Thelaxes suberi).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98797;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22077640; PubMed=12082135;
RA Fares M.A., Barrio E., Sabater-Munoz B., Moya A.;
RT "The evolution of the heat-shock protein GroEL from Buchnera, the
RT primary endosymbiont of aphids, is governed by positive selection.";
RL Mol. Biol. Evol. 19:1162-1170(2002).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ439085; CAD27797.1; -;
HAMAP; MF 00600; -; 1.
InterPro; IPR001844; Chaprinin Cpn60.
InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; IPR00298; CHAPERONIN60.
DR Pfam; PF00118; cpn60 TCP1; 1.

DR PRINTS; PRO0298; CHAPERONIN60.
DR PRINTS; PRO0304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 550 AA; 57787 MW; 100D3B923DA48502 CRC64;

Query Match 8.0%; Score 85; DB 1; Length 550;
Best Local Similarity 21.7%; Pred. No. 18;
Matches 49; Conservative 42; Mismatches 87; Indels 48; Gaps 10;

QY 13 VAVSDPPIHYDKITEINKAIDDA-IAALEQSETIDPMKVPDPHADKFERHVGIVDPKGL 71
DB 107 VAGGNPM-----DLKGVDKAVIAVEE---LKLSVPCSDPKAIAQVGTISANDS 156

QY 72 AMRNIEARGLKQKQGDANVKGEIGIVKAHLIIQVHDD---IVSMYDYLAKGLD--H 126
DB 157 TVGKLIATAEMKEV-----GKEGVITVEGTGLQDELVDVEGMDQDRGVLSYFINK 207

QY 127 PTHVISDIQDFVVALSLEISDE-----GNITWTSFVRQFA---VNVNHIG 170
DB 208 PTGSIELDNPILLADKKISNNREMLPPLSVAKAGNLLIIAEDVGEALATLVVNTMR 267

QY 171 GL-----SILDPIFGVLSVLTALFQDVTVRKEMTKVLAPAFKRELEK 212
DB 268 GIVKVAANKAPGFG---DRKAMLDIATLTGSTVISSEIGLELEK 310

RESULT 25
CH60_SERRU
ID CH60_SERRU STANDARD; PRT; 539 AA.
AC O66202;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOGA.
OS Serratia rubidaea.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=61652;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 1240;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AB008143; BAA25219.1; -;
HAMAP; P06139; 1GRL.
DR HAMAP; MF 00600; -; 1.
InterPro; IPR001844; Chaprinin Cpn60.
InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60 TCP1; 1.
DR PRINTS; PRO0298; CHAPERONIN60.
DR PRINTS; PRO0304; TCOMPLEXTCP1.

KW Chaperone; ATP-binding.
FT NON_TER 540 540
SQ SEQUENCE 540 AA; 56361 MW; 6DB51337F5C30D8C CRC64;
Query Match 7.9%; Score 84.5; DB 1; Length 540;
Best Local Similarity 18.6%; Pred. No. 19;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;
QY 13 VAVSADPTHYDKITEINKAIDDA-IAAEQSEITIDPMKVPDHDADKFERHVGIVDFKGL 71
DB 107 VAAGNPNM-----DLKRGIDKAVAAVEELKT---LSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIARGLKMKQGDANVKGEIVKAHLIIQVHDD---IVSMYEDLAY----- 120
DB 157 TVGKLIAMDKV-----GKEGVITVEDGTGLEDELVDVEGMDRGLYSPYFINK 207
QY 121 -----KLGDLHPTTHVVISDIQDFVVALSLEISDEGNITMTSFEV 159
DB 208 PETGAVELESPIILLADKKVSNIREMLPVLAVAKAGKPLVIAEDVEGEALATL----- 262
QY 160 RQFANVNHIGL-----SILDPFGVLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTGGTVISSEIGMELEK 311
RESULT 27
CH60_KLEOX STANDARD; PRT; 540 AA.
ID CH60_KLEOX STANDARD; PRT; 540 AA.
AC O66210;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
DE GSOL OR MOPA OR GROEL.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1665;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC EMBL; AB008147; BAA25227.1; -
CC HSSP; P06139; IJON.
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.

DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
FT NON_TER 539 539
SQ SEQUENCE 539 AA; 56640 MW; CE447B657515D4F0 CRC64;
Query Match 7.9%; Score 84.5; DB 1; Length 539;
Best Local Similarity 18.6%; Pred. No. 19;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;
QY 13 VAVSADPTHYDKITEINKAIDDA-IAAEQSEITIDPMKVPDHDADKFERHVGIVDFKGL 71
DB 107 VAAGNPNM-----DLKRGIDKAVAAVEE---LKKLSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNIARGLKMKQGDANVKGEIVKAHLIIQVHDD---IVSMYEDLAY----- 120
DB 157 TVGKLIAMDKV-----GKEGVITVEDGTGLEDELVDVEGMDRGLYSPYFINK 207
QY 121 -----KLGDLHPTTHVVISDIQDFVVALSLEISDEGNITMTSFEV 159
DB 208 PETGAVELESPIILLADKKVSNIREMLPVLAVAKAGKPLVIAEDVEGEALATL----- 262
QY 160 RQFANVNHIGL-----SILDPFGVLSDLVLTAFQDVTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISSEIGMELEK 311
RESULT 26
CH60_KLEOR STANDARD; PRT; 540 AA.
ID CH60_KLEOR STANDARD; PRT; 540 AA.
AC O66214;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
DE GSOL OR MOPA OR GROEL.
OS Klebsiella ornithinolytica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Raoultella.
OX NCBI_TaxID=54291;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 6096;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC EMBL; AB008149; BAA25231.1; -
CC HSSP; P06139; IJON.
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.

```

FT NON TER 540
SQ SEQUENCE 540 AA; 56445 MW; BFE441BC6B05E06B CRC64;

Query Match
Best Local Similarity 18.6%; Score 84.5; DB 1; Length 540;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPHYDKITEENKAIDA--TAAIQSTIDPMKVPDHPADKFEHVGIVDFKQEL 71
Db 107 VAGMFW-----DLKRGIDKAVAAVEELK--LSVFCSDSKAIAQVGRISANSDB 156
QY 72 AMRNIARGLKMQRGDANVKGEEGIVKAHLIGVHDD---IVSMEYDLAY----- 120
Db 157 TVGKLIAEAMDKV-----GKEGVITVDGTGLEDELVDVEGQFDRGYLSPYFINK 207

QY 121 -----KLGDLHPHTHVLSDIQDFVALSLSISDEGNITWTSFEV 159
Db 208 PETGAVELESFILLADKKSINREMLPVLVAKAKGKPLVIAEVDGEALATL----- 262

QY 160 RQFANVNHIGGL-----SILDPIFGVLSVLTAFQDVTVRKEMTKVLAFAFKRELEK 212
Db 263 -----VNTVRGIVKVAVKAPGFG---DRKAMLQDIATLTGGTVISBEIGNELEK 311

RESULT 28
RPB1 DROME
ID RPB1 DROME STANDARD; PRT; 1887 AA.
AC P04052; Q9VYX6;
DT 01-NOV-1986 (Rel. 03, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).
GN RPII215 OR Cg1554.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89218930; PubMed=2496296;
RA Jucker R.S., Weeks J.R., Zehring W.A., Greenleaf A.L.;
RT "Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila."
RL Mol. Gen. Genet. 215:266-275 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkely;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bock J., Brokstein P., Brotter P.,
RA Burtis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Hewland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirski R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [3]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE=87089662; PubMed=3025586;
RA Searles L.L., Greenleaf A.L., Kemp W.E., Voelker R.A.;
RT "Sites of P element insertion and structures of P element deletions in the 5' region of Drosophila melanogaster RpiI215."
RL Mol. Cell. Biol. 6:3312-3319 (1986).
RN [4]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=85282618; PubMed=2992806;
RA Biggs J., Searles L.L., Greenleaf A.L.;
RT "Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II."
RL Cell 42:611-621 (1985).
RN [5]
RP SEQUENCE OF 1441-1887 FROM N.A.
RX MEDLINE=88094402; PubMed=3122024;
RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function."
RL Mol. Cell. Biol. 8:321-329 (1988).
CC -! FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC -! CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).
CC -! SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO TWELVE DIFFERENT POLYPEPTIDES. THE 215 kDa POLYPEPTIDE IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
CC -! SUBCELLULAR LOCATION: Nuclear.
CC -! PTM: The tandem 7 residues repeats can be highly phosphorylated. The phosphorylation activates POL2.
CC -! MISCELLANEOUS: Three distinct zinc-containing RNA polymerases were found in eukaryotic nuclei: polymerase I for the ribosomal RNA precursor, polymerase II for the mRNA precursor, and polymerase III for SS and tRNA genes.
CC -! SIMILARITY: Belongs to the RNA polymerase beta' chain family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27431; AAA28968.1; --
CC EMBL; AE003486; AAF48057.1; --
CC EMBL; M14203; AAA28864.1; --
CC EMBL; M11798; AAA28863.1; --
CC EMBL; M19537; AAA28827.1; --
CC PIR; S04457; RNFF2L.
CC FlyBase; FBgn0003277; RpiI215.
CC InterPro; IPR000722; RNA_pol_A.
CC InterPro; IPR007080; RNA_pol_Rpb1_1.
CC InterPro; IPR007066; RNA_pol_Rpb1_3.
CC InterPro; IPR007083; RNA_pol_Rpb1_4.
CC InterPro; IPR007081; RNA_pol_Rpb1_5.

```

```
DR InterPro; IPR007075; RNA_pol_Rpb1_6.
DR InterPro; IPR007073; RNA_pol_Rpb1_7.
DR InterPro; IPR006592; RNA_pol_Rpb1_N.
DR InterPro; IPR000684; RNA_polIT_repeat.
DR Pfam; PF04997; RNA_pol_Rpb1_1.
DR Pfam; PF00623; RNA_pol_Rpb1_2.
DR Pfam; PF04983; RNA_pol_Rpb1_3.
DR Pfam; PF05000; RNA_pol_Rpb1_4.
DR Pfam; PF04998; RNA_pol_Rpb1_5.
DR Pfam; PF04992; RNA_pol_Rpb1_6.
DR Pfam; PF04990; RNA_pol_Rpb1_7.
DR Pfam; PF05001; RNA_pol_Rpb1_R.
DR SMART; SM00663; RPOA1_N; 1.
DR PROSITE; PS00115; RNA_POL_II_REPEAT; 11.
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 67 83 C2H2-TYPE (POTENTIAL).
FT DOMAIN 1579 1881 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT CONFLICT 319 324 RAMQKS -> GYAKV (IN REF. 4).
FT CONFLICT 450 450 F -> G (IN REF. 4).
FT CONFLICT 455 458 TLHK -> RCTT (IN REF. 4).
FT CONFLICT 463 472 GHRVKVLPWS -> VTGESVASST (IN REF. 4).
FT CONFLICT 741 741 R -> H (IN REF. 1).
FT CONFLICT 1485 1524 SMLGGAAMFVGSGTSPMTPTWANCNTPRYFSPGHV
-> I (IN REF. 5).
FT CONFLICT 1506 1508 MTP -> ELDSA (IN REF. 1).
FT CONFLICT 1887 1887 D -> DVKKGGRG (IN REF. 1).
FT SEQUENCE 1887 AA; 209167 MW; 4EC68C7708A167A3 CRC64;

Query Match 7.9%; Score 84.5; DB 1; Length 1887;
Best Local Similarity 22.7%; Pred. No. 82;
Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;

QY 17 ADPIHYDKITEINKAIDDAATAAEQSETIDPMKVPDHA--DKERHVGIV-----DFKG 69
DB 686 ADPQTYNEIQAIKAKDDVINIQAHNMELETPGNTLRQTFENVKNVILNDARDKTG 745
QY 70 ELAMENI-EARGLKQMKRQ-----GDANKGEE---GIVKAHLILIGVHD 109
DB 746 GSAKSLTEYNNLKAMVYSGSGKSNINISQIVACGQGVNVEGKIPYGRKTLPHFIKD 805
QY 110 D-----IVSNEVDLAYKLGDLHPTT---HVISDIQDFVVALSLEISDSEGN-----T 133
DB 806 DYGPEISRGFVENS-----LAGLPSEFYFHAMGG-REGLIDTAVKTAETGYIQRLLIKA 859
QY 154 MTSFEVROFANVNHIGLSLIL---DPIFGVLSID----- 184
DB 860 MESVNVNDGTIVRSVGGILQIRYGEDGLGELVEFQNPVTKLSNKSFEKRFKFDWSNE 919
QY 185 -VLTAFQDTYRKEMT 199
DB 920 RLMKKVFTDDVIKEMT 935

RESULT 29
SYFA STRCO
ID SYFA STRCO STANDARD; PRT; 358 AA.
AC O88055;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)
DE (Phenylalanyl-tRNA ligase alpha chain) (PHERS).
GN PHES OR SC01595 OR SC195.17C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=215996410; PubMed=12000953;
RA Bentley S.D., Chater K.P., Cerdano-Tarraga A.-M., Challis G.L.,
```

```
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
Hopwood D.A.;
"Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
Nature 417:141-147(2002).
RL NATURE 417:141-147(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
Phe-tRNA synthetase alpha chain subfamily 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch)
-----
DR EMBL; AL939109; CNA20806.1; ALT_INIT.
DR HSSP; P27001; LPYS.
DR HAMAP; MF_00281; -.
DR InterPro; IPR004188; Phe tRNA_synth_N.
DR InterPro; IPR004529; PheS.
DR InterPro; IPR002319; tRNA-synth_2d.
DR InterPro; IPR006195; tRNA_ligase_II.
DR Pfam; PF02912; Phe-tRNA-synth_N; 1.
DR Pfam; PF01409; tRNA-synth_2d; 1.
DR TIGRfam; TIGR00468; pheS; 1.
DR PROSITE; PS00862; AA tRNA LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
Metal-binding; Magnesium; Complete proteome.
KW METAL-BINDING; MAGNESIUM (BY SIMILARITY).
FT METAL 262 262 MAGNESIUM
SQ SEQUENCE 358 AA; 39119 MW; 49CF6E0000E469C CRC64;

Query Match 7.9%; Score 84; DB 1; Length 358;
Best Local Similarity 23.4%; Pred. No. 13;
Matches 33; Conservative 26; Mismatches 50; Indels 32; Gaps 5;

QY 27 EINKAIDDAIAAIQSEITIDPMKVPDHADK-----FERHVGIV--DEKGLAMENI 76
DB 4 EEIERMRDEALAAFAAADSLDALQAKVAHTGGASPLANREIGALPPQAKAEAGKRVG 63
QY 77 EARGL-----KQMRQGDANYKGEIGVKAHLIGVHDDIVSMEYDLYAYKLGDLHP 127
DB 64 MARGAVNKALAAQOELEGAERDARVLVEAV-----DVTLPHD-RVPAGARHP 110

RESULT 30
DEOB STRAM
ID DEOB STRAM STANDARD; PRT; 392 AA.
AC Q99X76;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phosphotransferase (EC 5.4.2.7) (Phosphoesterase)
GN DEOB OR DRM OR SAV0139 OR SA0134 OR MW0113.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
```

```
OC Bacteria: Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MUSO / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Ohshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwano M., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827 (2002).
CC -!- FUNCTION: Phosphotransfer between the C1 and C5 carbon atoms of
CC pentose (By similarity).
CC -!- CATALYTIC ACTIVITY: D-ribose 1-phosphate = D-ribose 5-phosphate.
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 1-phosphate = 2-deoxy-D-
CC ribose 5-phosphate.
CC -!- COFACTOR: Binds 1 or 2 manganese ions (Potential).
CC -!- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the phosphotransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003358; BAB56301.1; -.
CC EMBL; AP003129; BAB41354.1; -.
CC EMBL; AP004822; BAB93978.1; -.
CC PIR; G89774; G89774.
CC SWISS-2DPAGE; O99X76; STRAN.
CC HAMAP; MF_00740; -.
CC InterPro; IPR006124; Metalloenzyme.
CC Pfam; PF01676; Metalloenzyme; 1.
CC TIGRFAMs; TIGR01696; deob; 1.
CC Isomerase; Metal-binding; Manganese; Complete proteome.
CC METAL 14 14 MANGANESE (BY SIMILARITY).
CC METAL 291 291 MANGANESE (BY SIMILARITY).
CC METAL 327 327 MANGANESE (BY SIMILARITY).
CC METAL 328 328 MANGANESE (BY SIMILARITY).
CC METAL 339 339 MANGANESE (BY SIMILARITY).
CC SEQUENCE 392 AA; 43795 MW; 6012D71CD9BDB549 CRC64;
Query Match 7.9%; Score 84; DB 1; Length 392;
Best Local Similarity 24.1%; Pred. No. 15;
Matches 45; Conservative 35; Mismatches 85; Indels 22; Gaps 10;
QY 14 AVSADPHYKITE-EINKAIDDAIAAISOSETIDPMKVPDPAKFERHVGIVDFKGLA 72
DB 65 AVEQPEAYTTLSEASYGKDTGTGHEWIMGLNIMQFPKVPNGFPBELLQIQBEMTGRKV 124
QY 73 MRNIEARGLKMRQGDANVKGEIGVKAH---LLIGVHDDIVSME--YDLAYKLGDL- 125
DB 125 VANKPASGTQIIEWGSHQKMTGDLIVTSADPVLQIAHEDIIPLELYDICKERELT 184
QY 126 HPTHVTSIDIQFVVALSLEISDEGNITMTS-----FEVROFA-NVNVHI--GGLSILDPI 178
DB 185 KDPKYLIGRI-----IARPY-VGEPGNFTRTSNRHDYALKPFGKTYLDHLKDDGYDVI--A 237
QY 179 FGVLSDV 185
DB 238 IGKINDI 244
RESULT 31
XYLG ECOLI STANDARD; PRT; 513 AA.
AC P37388;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE D-xylose transport ATP-binding protein xylG.
GN XYLG OR B3567.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586 (1994).
CC -!- FUNCTION: INVOLVED IN THE HIGH-AFFINITY D-XYLOSE MEMBRANE
CC TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U00039; AAB18544.1; -.
CC EMBL; AE000434; AAC76591.1; -.
CC PIR; S47788; S47788.
CC HSSP; Q58663; 1G6H.
CC EcoGene; EG12275; xylG.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 2.
CC ProDom; PD000006; ABC_transporter; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
CC Transport; Sugar transport; Inner membrane; ATP-binding;
CC Complete proteome.
CC NP_BIND 37 44 ATP (POTENTIAL).
CC SEQUENCE 513 AA; 56470 MW; FAD2C65D178D50DF CRC64;
Query Match 7.9%; Score 84; DB 1; Length 513;
Best Local Similarity 23.5%; Pred. No. 20;
Matches 44; Conservative 32; Mismatches 71; Indels 40; Gaps 7;
QY 31 KAIDDAIAAISQSETID-----PMKVPDPAKFERHVGIVDFKGLAMRNIEAR 79
DB 18 KAIDNVCLRNAGSIVSLGCGSGKSTLMKVLGCIYPHGSYEGEIIFAGE-----EIQAS 73
QY 80 GLKQMRQGDANVKGEIGVKAH---AHLIG---VHDDIVSMEYDLYAYKLGDLHPHTV 131
DB 74 HIRTERGKIAIIHQELALVKELTVLENIPLGNIEITNGI--MDYDL-----M 119
QY 132 ISDIQDFVVALSLEISDE---GNITMTSFEVROFANVNVHIGGSLDPIFGVLSDVLT 188
```

```
Db 120 TLRCQKLAQVSLSPTRVGLGQQLVETAKALNKQVRLILDEPTASLITEQETS 179
QY 189 IFQDTR 195
Db 180 ILDIIR 186

RESULT 32
CH60_ENTGE STANDARD; PRT; 540 AA.
AC O66194;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin 1 (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOPA.
OS Enterobacter gergoviae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=61647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1234;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008139; BAA25211.1; -
CC HSSP; P06139; IJON.
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaprln Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; Cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC NON TER 540 540
CC SEQUENCE 540 AA; 56617 MW; B6DA54F0F5F84BBD CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 540;
Best Local Similarity 18.2%; Pred. No. 23;
Matches 43; Conservative 41; Mismatches 85; Indels 67; Gaps 8;

QY 13 VAVSADPHYDKITBEINKAIDDAIAEQSETIDPMKVPDPHADKFERHVGIVDFKGEIA 72
Db 107 VAGNPNM-----DLKRGIDKAVAA--AVEELKALSVPCSDSKATQAQVGTISANSDET 157
QY 73 MENIARGLKQMKRGDANVGEIGVKAHLILGVHDD---IVSMEDYLA----- 120
Db 158 VGKLTAEMARKV-----GKEGVITVEDGTGLQDELVDVEGQDRGVLSPYFINKP 208
QY 121 -----KLGLDHPHTTHVISDIQDFVWALSLEISDEGNITWTSFEVR 160
```

```
Db 209 ETGAVELSDPFFILLADKKISNIRELLPVLVAVAKAGKPELLIIAEDVEGEALTL----- 262
QY 161 QPANNVNHGGI-----SILDFIFGVLSVLTAFQDTRKEMTKVLAPAFKRELEK 212
Db 263 -----VVNTMRGIVKVAVKAPGFG--DRRKAMLDQDIATLTAGTVISSEIGMELEK 311

RESULT 33
CH61_VIBVU STANDARD; PRT; 546 AA.
AC O9ALA9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 60 kDa chaperonin 1 (Protein Cpn60) (groEL protein 1).
GN GROEL OR GROEL OR GROEL OR VWI1260.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RA Wong H.-C., Lu K.-H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Khee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.B.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY017169; RAG48876.1; -
CC HSSP; P06139; IGRU.
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaprln Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; Cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding; Complete proteome.
CC FT CONFLICT 134 134 L -> M (IN REF. 1).
CC FT CONFLICT 220 220 I -> F (IN REF. 1).
CC FT CONFLICT 256 256 G -> D (IN REF. 1).
CC FT CONFLICT 272 272 K -> I (IN REF. 1).
CC FT CONFLICT 277 272 K -> I (IN REF. 1).
CC FT CONFLICT 388 388 E -> K (IN REF. 1).
CC FT CONFLICT 474 474 G -> A (IN REF. 1).
CC FT CONFLICT 545 545 M -> GM (IN REF. 1).
CC SQ SEQUENCE 546 AA; 57515 MW; 92869B7E517B3D2A CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 546;
Best Local Similarity 25.0%; Pred. No. 23;
Matches 42; Conservative 19; Mismatches 70; Indels 37; Gaps 5;
```


QY 6 LIAAVAFVAVSA-----DPIHYDKITEINKAIDDAATAAEQSETIDPMKV 51
Db 372 LAGGAVIKVGAATEVEMKEKORVEDALHATFAAVEGIVAGGVGVALIRAASKI-----V 427
QY 52 PDHADKFERHVGIVDFKGLAMRNIEARGLKQMKROGDA-----NVKGEIGIVKAHLI 105
Db 428 DLOGDNEEQNGVI-----RVALRAMEAPLRQITKNAGDEESVANNVRAGEGSGYNAAT 482
QY 106 GHVDDVSMVEYDLAYKGLDHPHTTHVISIDQFVVALSLEISDEGNIT 153
Db 483 GVGVDMLEM-----GLDPTKVTFSALQPAASVAGLMITAEAVT 522

RESULT 34
CH60_YERPE STANDARD; PRT; 548 AA.
AC Q821Y3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL OR MOPA OR YPO0351 OR Y0609.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Farraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holtroyd S., Jagels K., Karyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ414142; CAC89210.1; -;
DR EMBL; AE013663; AAM84197.1; -;
DR PIR; AG0043; AG0043.
DR HAMAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaperonin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR InterPro; IPR008950; GroEL-ATPase.

DR Pfam; PF00118; cpn60_TCF1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCPI.
DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 548 AA; 57431 MW; 6E33B1F889B79E3A CRC64;

Query Match 7.8%; Score 83.5; DB 1; Length 548;
Best Local Similarity 18.1%; Pred. No. 23;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VASADPIHYDKITEINKAIDDA-IAAIFQSETIDPMKVPDHDADKFERHVGIVDFKGL 71
Db 107 VAAGNPM-----DLKRGIDKAVIAAVEE---LKLSPVCSKRAIAQVGTISANDS 156
QY 72 AMRNEARGLKQMKROGDANVKGEIGIVKAHLIGVHDD---IVSMVEYDLAY----- 120
Db 157 TVGELIAQAMEKV-----GKEGVITVEEGSLQDELVDVGEQMGDFRGYLSPYFINK 207
QY 121 -----KLGDLHPTTHVISIDQFVVALSLEISDEGNITMTSFEV 159
Db 208 PETGSIIELESPFILLADKKISINIREMLPVLEAVAKAGKPLLIITAEVGEALATL----- 262
QY 160 RQFANVNHIGGL-----SILDPFPGVLSDLVITAFQDTVRKENTKVLAPAFKRELEK 212
Db 263 -----VNTMRGIVKVAAPGFG---DRKAMQLQDIATLTAGTVISEIGLELEK 311

RESULT 35
CH60_YEREN STANDARD; PRT; 550 AA.
AC P49219;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein) (Heat shock protein
DE 60) (Cross-reacting protein antigen).
GN GROEL OR GROEL OR MOPA OR HSP60 OR CRPA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y-108C / Serotype O:3;
RA Hafner C.E., Roggenkamp A.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA / Serotype O:8;
RA Autenrieth I.B., Noll A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype O:3;
RX MEDLINE=94248434; PubMed=8190995;
RA Yamamoto T., Miura H., Ohsumi K., Yamaguchi H., Taguchi H.,
RA Ogata S.;
RT "Cloning and nucleotide sequence analysis of immunodominant
RT heat-shock protein of Yersinia enterocolitica."
RL Res. Microbiol. 144:1691-701(1993).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ414142; CAC89210.1; -;
DR EMBL; AE013663; AAM84197.1; -;
DR PIR; AG0043; AG0043.
DR HAMAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaperonin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR InterPro; IPR008950; GroEL-ATPase.


```

SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM 1-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -|- COFACTOR: Zinc (By similarity).
CC -|- SIMILARITY: Belongs to peptidase family M18.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE007625; AAK79065.1; -.
DR PIR: F97034; F97034.
DR HAMAP: MF_00466; -.
DR InterPro: IPR001948; Peptidase_M18.
DR Pfam: PF02127; Peptidase_M18; 1.
DR PRINTS: PR00932; AMINOPEPTASE.
KW Hypothetical protein; Hydrolase; Amino peptidase; Metalloprotease;
KW Zinc; Complete proteome.
FT METAL 105 105 ZINC (POTENTIAL) .
FT METAL 180 180 ZINC (POTENTIAL) .
FT METAL 441 441 ZINC (POTENTIAL) .
SQ SEQUENCE 465 AA; 52011 MW; CB4C6D0A54C0A439 CRC64;

Query Match 7.8%; Score 83; DB 1; Length 465;
Best Local Similarity 28.4%; Pred.No. 21; 49; Indels 30; Gaps 10;
Matches 4; Conservative 27; Mismatches

QY 74 RNIE-- --ARGLKQMKRQGD---ANVK-----GEGIVKAAHLITGVHDDI--VSMEY 116
DB 58 RNIEDILAKG-ETLKEGDKVYANNRGKGLIMFLIGKEPLYTGFKILGAHIDSPRLDLKQ 115
QY 117 DLAYKLGDLHP-TTHVLSIDTQDFV-VALSLEISDEGNITWTSFVROFANVNVHIGLSI 174
DB 116 NPLYEDTDLAMLETHYYGGIKYKKQWVTLPLAI--HGV-----VKKDGTIVNVCVGEDD 167
QY 175 LDPIFGVLSDLTAIFODTVRKEMTKVL 202
DB 168 NDFVFGV-SDILVHLASEQLEKASKVI 194

RESULT 38
PUR9 SYNEL STANDARD; PRT; 518 AA.
AC Q8DI9;
CT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bifunctional purine biosynthesis protein purH [Includes:
DE Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
DE (AICAR transformylase); IMP cyclohydrolase (EC 3.5.4.10) (inosinase)
DE (IMP synthetase) (ATIC)].
GN PURH OR TLR1547
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22235144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Katayabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,

```

Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
"Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
DNA Res. 9:123-130(2002).
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5-amino-1-(5-
phospho-D-riboseyl)imidazole-4-carboxamide = tetrahydrofolate + 5-
formamido-1-(5-phospho-D-riboseyl)imidazole-4-carboxamide.
CC -1- CATALYTIC ACTIVITY: IMP + H(2)O = 5-formamido-1-(5-
phosphoribosyl)imidazole-4-carboxamide.
CC -1- PATHWAY: De novo purine biosynthesis; ninth step.
CC -1- PATHWAY: De novo purine biosynthesis; tenth step.
CC -1- DOMAIN: The IMP cyclodehydratase activity resides in the N-terminal
region (by similarity).
CC -1- SIMILARITY: Belongs to the purH family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; AP005374; BAC09099.1; -.
CC DR HAMAP; MF 00139; -. 1.
CC DR InterPro; IPR002695; AICARFT_IMPChas.
CC DR InterPro; IPR004362; MGS like.
CC DR Pfam; PF01808; AICARFT_IMPChas; 1.
CC DR Pfam; PF02142; MGS; 1.
CC DR ProDom; PD004666; AICARFT_IMPChas; 1.
CC DR TricFRAMS; TIGR00355; purH; 1.
CC DR Purine biosynthesis; Transferase; Hydrolase; Multifunctional enzyme;
KW Complete proteome.
CC
CC SEQUENCE 518 AA; 55271 MW; F1010495BDCFE6 CRC64;
Query Match 7.8%; Score 83; DB 1; Length 518;
Best Local Similarity 23.5%; Pred.No. 24;
Matches 52; Conservative 34; Mismatches 89; Indels 46; Gaps 10;
QY 11 AFVAVSADPHYDKITEINKAIDDAIALEQSETIDPMKVPDADKFERHVGIVDFKGE 70
DB 140 AHVTLVDSQVETYLQEL-RLYGDQPAPFLACAAQAFALTAYDQ-----AIAYLQK 193
QY 71 LAMRNTEARGL-----KMKRQGDA-----NVKG--EEGIVKAHLIG---VH 108
DB 194 VTAAGTEPELTPPVFHLTGKQKQVLYAGENPHQRASWVCGAHPRGWAAAHLLQKLSY 253
QY 109 DDIVSWEYD---LAYKLGDLPHPTHTVSDIQDFVALSLEISDEGNITMTSFEVRFANV 165
DB 254 NLLLDLEARGVISEFLGSDSPAIVIKHTNPGVA-----EGKTLVEAYERFAADR 306
QY 166 VNHIGLSILD-PIFGVLSDLVLTAFQDVTVRKEMTKVLAPA 205
DB 307 VSAFGGIVALNRPLDGATAEALTRTFLEC-----VVAFA 340
RESULT 39
CH60_BACFO
ID CH60_BACFO STANDARD; PRT; 543 AA.
AC P81284; Q9X6Y3;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaparonin (Protein Cpn60) (groEL protein).
GN GROL OR GROL OR MOPA.
OS Bacteroides forsythus.
OC Bacteroidetes; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Tannerella.
OX NCBI_TaxID=28112;
RN [1]
RP
RC SEQUENCE FROM N.A.
RC STRAIN=ATCC 43037;
RX MEDLINE=99452396; PubMed=10524765;

Best Local Similarity 21.2%; Pred. No. 55;
Matches 32; Conservative 29; Mismatches 38; Indels 52; Gaps 8

QY 91 NVKGE-----EGIVKAHLLIGVDDIVSMEDVLAYK-----LGDLPHTTHVTSDIQDF-- 138
 | | | | | | | |
Db 6 NTKTEYSFLDSWVKV-----DDYLETAHLRGYQTVGICDVGNLHAAPFRVKQAQKNL 58
 | | | | | | | |
QY 139 --VVALLSLEISDEG-----N1-----TMTSFVFQFANYVNHHIGLSILD 176
 | | | | | | | |
Db 59 QPIISIELNFWRGPUGTAFSFIADTEGYKNILIRISTLNHYGERQSPDIQNHLSGIALII 118
 | | | | | | | |
QY 177 P-IFGVSLDVLTAIPQDTVRKEMTKYLAPAF 206
 | | | | | | | |
Db 119 PTYTGSLSEL-----TESSSVADEAF 139
 | | | | | | | |

```

RESULT 42
CBIO_METJUA STANDARD; PRT; 279 AA.
AC Q58486;
AD ID - CBIO_METJUA STANDARD; PRT; 279 AA.
DT DT 10-OCT-2003 (Rel. 42, Created)
DT DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable cobalt transport ATP-binding protein cbio.
GN CBIO OR MJI088;
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcaceales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Rutt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sult G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervilave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glöckle A.,
RA Scott J.L., Geoghegan N.S.M., Peterson J.F., Fuhrman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Weidman J.D., Sadow P.W., Hanna M.C.,
RA Clifton H.P., Roberts K.M., Hurst M.O., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).
CC -!- FUNCTION: May be involved with cobalt transport in association
CC with cobalamin biosynthesis (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licens@isb-sib.ch).
CC -----
DR EMBL; U67551; AAC99089.1; --
DR PIR; G64435; G64435.
DR HSPP; OS8663; IG6H.
DR TIGR; MJ1088;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005876; CblN_ATP.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMS; TIGR01166; cbio; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Cobalamin biosynthesis; Cobalt transport; Transport; ATP-binding;
KW Complete proteome.
FT NP BIND 37 43
SQ SEQUENCE 279 AA; 31120 MW; 7A2DF31BE6SCFA81 CRC64;

```

[illegible]

```

RESULT 43
CH60_ENTAM
ID CH60_ENTAM STANDARD; PRT; 539 AA.
AC O66196;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
DE GROEL OR GROEL OR MOFA.
OS Enterobacter amnigenus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=61646;
[1]
SEQUENCE FROM N.A.
RN
RP STRAIN=JCM 1237;
RC
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RT J. Gen. Appl. Microbiol. 43:355-361(1997).
RL
RC -! FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC
CC -! SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC
CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC -! SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL; AB008140; BAA25213.1; --
CC
CC HSSP; P06139; 1JON.
CC
CC HAMAP; MF_00600; --; 1.
CC
CC InterPro; IPR001844; Chapernin Cpn60.
CC
CC InterPro; IPR002423; Cpn60/TCP-1.
CC
CC InterPro; IPR008950; GroEL-ATPase.
CC
CC Pfam; PF00118; cpn60_TCP1; 1.
CC
CC PRINTS; PR00239; CHAPERONIN60.
CC
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC
CC Chaperone; ATP-binding.
CC
CC NCN_TER 539 539
CC
CC SEQUENCE 539 AA; 56657 MW; 09E701CB2B1DBA17 CRC64;
CC
CC FT
CC

```

```

Query Match          7.7%; Score 82.5; DB 1; Length 539;
Best Local Similarity 18.6%; Pred. No. 28;
Matches 44; Conservative 43; Mismatches 81; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAALEQSEITDPMKVPDHADYFERHVGIVDFKGL 71
DB 107 VVAGNNPM-----DLKRGDKAVTAAVELKT---LSVPCSDSKAIAQVGTISANSDE 156
QY 72 AMRNTBARGLKQMKRGDANVKGEGIVKAHLIGVHD---IVSNEYDLAY-----120
DB 157 TVGKLIABEMDKV-----GKEGVITVEDGTGLEDELVDVVEGQDFRGYLSFYFINK 207
QY 121 -----KLGDLPHTTHVISDIQDFVWALSLSISDEGNITMTSFEV 159
DB 208 PETGAVELESPFILLADKKISINIREMLFVLEAVAKAGKELLIAEDVEGEALATL---262
QY 160 RQFANYVNHIGGL-----SILDPIFGVLSDLVTAIFQDTPVRKEMTKYLAPFAFKRELEK 212
DB 263 -----VVNTMRGIVKVAAPKAPGFG---DRRKAMLDQIATLTGGTVISEIGMELEK 311

RESULT 44
CH60_KLEPL STANDARD; PRT; 540 AA.
AC O66212;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
OS GROS OL MOFA OR GROEL.
OS Klebsiella planticola (Raoultella planticola).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Raoultella.
OX NCBI_TaxID:575;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 7251;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008148; BAA25229.1; -.
CC HSP; P06139; IJON.
CC HAMAP; MF 00600; -.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC DR Chaperone; ATP-binding.
CC NON_TER 540 540
CC FT SEQUENCE 540 AA; 56655 MW; 9E118C11ABC2430A CRC64;

Query Match          7.7%; Score 82.5; DB 1; Length 540;

```

	Matches	43;	Conservative	44;	Mismatches	81;	Indels	69;	Gaps	9;
Qy	13	YVASADPIHYDKITEINKA	DDA--TAATQ	STSIDPMKVPD	HADK	EERHVG	IVDFK	GEL	71	
		YVASADPIHYDKITEINKA	DDA--TAATQ	STSIDPMKVPD	HADK	EERHVG	IVDFK	GEL	71	
Db	107	VAAAGMNM-----	DLKGI	DKAVVA	VEE---LKKLS	VPCSDSKAIAQ	VG	TIANSDE	156	
		VAAAGMNM-----	DLKGI	DKAVVA	VEE---LKKLS	VPCSDSKAIAQ	VG	TIANSDE	156	
Qy	72	AMENIEARGIKOMKROGD	ANVK	BEGEIVK	AHLLIG	VHDD--I	VSMEYD	IAY	120	
		AMENIEARGIKOMKROGD	ANVK	BEGEIVK	AHLLIG	VHDD--I	VSMEYD	IAY	120	
Db	157	TVGKLIAEAMEKV-----	KGEG	VITVEEGTGL	QDEL	DVV	EGM	QFDRG	YLS	207
		TVGKLIAEAMEKV-----	KGEG	VITVEEGTGL	QDEL	DVV	EGM	QFDRG	YLS	207
Qy	121	-----	KLGD	LPHTHV	ISD	IQDF	VVAL	SLEIS	DEGNI	159
		-----	KLGD	LPHTHV	ISD	IQDF	VVAL	SLEIS	DEGNI	159
Db	208	PETGSVELSPFILLADK	KSIN	TRELLP	VL	EA	VAK	GPKLLI	VAED	262
		PETGSVELSPFILLADK	KSIN	TRELLP	VL	EA	VAK	GPKLLI	VAED	262
Qy	160	RQFANVNNHIGGL-----	SILD	PTFGVL	SDVL	TAI	PD	TV	RKEMTK	212
		RQFANVNNHIGGL-----	SILD	PTFGVL	SDVL	TAI	PD	TV	RKEMTK	212
Db	263	-----	VVNT	NRGIVK	VA	AVK	APG	---DRK	AMLQ	311
		-----	VVNT	NRGIVK	VA	AVK	APG	---DRK	AMLQ	311

```

CH60_XANAC
ID   CH60_XANAC          STANDARD;          PRT;      546 AA.
AC   Q9PFX1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaparonin (Protein Cpn60) (groEL protein).
GN   GROEL OR GROEL OR XAC0542.
OS   Xanthomonas axonopodis (pv. citri).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=92829;
[1]
SEQUENCE FROM N.A.
RP   STRAIN=306 / ATCC 13902 / XV 101;
RC   STRAIN=306 / ATCC 13902 / XV 101;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bartolini M.C., Camargo L.E.A.,
RA   Camarotte G., Camnava F., Cardozo J., Chambergo P., Cipina L.P.,
RA   Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Merck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.P., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC   proper assembly of unfolded polypeptides generated under stress
CC   conditions (By similarity).
CC   -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC   7 subunits (By similarity).
CC   -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -1- SIMILARITY: Belongs to the chaparonin (HSP60) family.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   ENBL; AE011680; AM35431.1; -.
DR   HAMAP; MF 00600; -. 1.
DR   InterPro; IPR001844; Chaprinin Cpn60.

```

DR HAMAP; MF_00600; -; 1.
DR InterPro; IPR001844; Chaprinin_Cpn60.

DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PRO0298; CHAPERONIN60.
DR PRINTS; PRO0304; TCOMPLEXCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 546 AA; 57166 MW; 07D167E4347D5C9 CRC64;

Query Match 7.7%; Score 82.5; DB 1; Length 546;
Best Local Similarity 20.8%; Pred. No. 28;
Matches 50; Conservative 40; Mismatches 79; Indels 71; Gaps 11;

QY 13 VAVSADPIHYDKITEENKAIIDAI-AAIEQSETIDPMKYPDHRADKPERHVGIVDFKGL 71
DB 107 VAAGNPM-----DLKRGIDQAKAAVVELKNISK---PTTDDKAAIAQVGTISANSD 156
QY 72 AMRNIIEARGLKMQRGDANVKGEGIVKAHLIGVH---DDIVSMEDYLAY----- 120
DB 157 SIGNIIIEAMKV-----GKEGIVTEEGSGLENELDVVEGMQFDRGYLSPYFINN 207
QY 121 -----KLGDLHPTTH--VISDIQDFVAL-----SLEISDGNITMTSFEV 159
DB 208 QQSQSADLDDPFILLHDKISNRVLDLPVLEGVAKAGKPLIIAAVEVEGEALATL----- 262
QY 160 QFANVNHIGGL-----SILDPIFG-----VLSDLTAIFQDTVRKEMTKVLAPEKREL 210
DB 263 -----VNTINGIVKVVAVKAPGFGDRRKAMLEDMVLTGTVISEEVLGALEKATIKOL 317

RESULT 47
ID COA2 HUMAN STANDARD; PRT; 2483 AA.
AC O00763; Q16852;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Acetyl-CoA carboxylase 2 [EC 6.4.1.2] (ACC-beta) [Includes: Biotin
carboxylase [EC 6.3.4.14]].
GN ACACB OR ACC2 OR ACCB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
[1]
RC TISSUE=Liver;
RC SEQUENCE FROM N.A.
RX MEDLINE=97256787; PubMed=9099716;
RA Abu-Elheiga L., Almaraz-Ortega D.B., Baldini A., Wakil S.J.;
RT "Human acetyl-CoA carboxylase 2. Molecular cloning, characterization,
RL J. Biol. Chem. 272:10669-10677 (1997)."
[2]
RC TISSUE=Adipose tissue;
RX MEDLINE=9626506; PubMed=8670171;
RA Widmer J., Fasshi K.S., Schlichter S.C., Wheeler K.S., Crute B.E.,
RA King N., Nuttle-Mcmenemy N., Noll W.W., Daniel S., Ha J., Kim K.-H.,
RA Witters L.A.;
RT "Identification of a second human acetyl-CoA carboxylase gene.";
RL Biochem. J. 316:915-922 (1996).
CC -!- FUNCTION: ACC-beta may be involved in the provision of malonyl-CoA
CC or in the regulation of fatty acid oxidation, rather than fatty
CC acid biosynthesis. This protein carries three functions: biotin
CC carboxyl carrier protein, biotin carboxylase, and
CC carboxyltransferase.
CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
CC + malonyl-CoA.
CC -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)
CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.
CC -!- COFACTOR: Biotin.
CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first (rate-limiting)
CC step.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC; MAY ASSOCIATE WITH MEMBRANES.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=O00763-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=O00763-2; Sequence=VSP_000547;
CC -!- TISSUE SPECIFICITY: Predominantly expressed in the heart, skeletal
CC muscles and liver.
CC -!- SIMILARITY: PARTIAL TO CARBAMOYL PHOSPHATE SYNTHETASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U89344; AA58382.1; -;
CC EMBL; U34591; AAC50571.1; -;
CC FIC; S71091; S71091.
CC HSSP; P24182; 1DV1.
CC Genew; HGNC:85; ACACB.
CC MIM; 601557; -;
CC InterPro; IPR001882; Biotin_BS
CC InterPro; IPR005482; Biotin_carb_C.
CC InterPro; IPR000889; Biotin_lipoyl.
CC InterPro; IPR000022; Carboxyl_trans.
CC InterPro; IPR005479; Cphase_L_D2.
CC InterPro; IPR005481; Cphase_L_N.
CC Pfam; PF02785; Biotin_carb_C; 1.
CC Pfam; PF00364; biotin_lipoyl; 1.
CC Pfam; PF00309; Carboxyl_trans; 1.
CC Pfam; PF00289; Cphase_L_D2; 1.
CC Pfam; PF02786; Cphase_L_D2; 1.
CC PROSITE; PS00188; BIOTIN; 1.
CC PROSITE; PS00866; CPSASE 1; 1.
CC PROSITE; PS00867; CPSASE 2; 1.
KW Fatty acid biosynthesis; Biotin; Ligase; Multifunctional enzyme;
KW ATP-binding; Phosphorylation; Alternative splicing.
FT NP_BIND 458 463
FT ACT_SITE 584 584
FT BINDING 927 927
FT DOMAIN 2095 2124
FT MOD_RES 219 219
FT MOD_RES 221 221
FT VARSPIC 1115 1215
FT CONFLICT 1365 1365
FT CONFLICT 1375 1375
FT CONFLICT 1408 1408
FT CONFLICT 1421 1421
FT CONFLICT 1450 1450
FT CONFLICT 1487 1487
FT CONFLICT 1555 1558
FT CONFLICT 1609 1609
FT CONFLICT 1630 1630
FT CONFLICT 1844 1844
FT CONFLICT 1917 1918
FT CONFLICT 2483 AA; 279690 MW; AAC678276B49F753 CRC64;
SQ SEQUENCE 2483 AA; 279690 MW; AAC678276B49F753 CRC64;

Query Match 7.7%; Score 82.5; DB 1; Length 2483;
Best Local Similarity 24.2%; Pred. No. 1.6e+02;
Matches 39; Conservative 32; Mismatches 49; Indels 41; Gaps 7;

QY 45 TIDPMKVPDHRADKPERHVGIVDFKGL-----AMRNIIEARGLKMQRGQ----- 88
DB 2235 TINPLCIEMWADK-ESRGVLEPGSTVEIKEREDLIKSMRRIDPAYKLMQELGEPDLS 2293
QY 89 DANVKGEGIVKAHLIGVHDDIVSMEDYLAYKUGDLHPT-----THVLSDDQDFVAL 142
DB 2294 DKDRKDLKLEGRUKAR-----EDLLPIYHQVAVQPADFHDTPGRMLEKGVISDLEWKTAR 2348


```
-!- SIMILARITY: Belongs to class-I pyridine nucleotide-disulfide  
oxidoreductase family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; AE001663; AAD18970.1; --  
EMBL; AE002261; AAP38812.1; --  
EMBL; AP002548; BRA99041.1; --  
EMBL; AB035943; BAA8651.1; --  
EMBL; AE017160; AAP98791.1; --  
PIR; C72031; C72031.  
PIR; G86594; G86594.  
HSSP; P14218; ILPF.  
PHCI-2DPAGE; Q92773; --  
TIGR; CP1037; --  
InterPro; IPR001327; FAD_pyr_redox.  
InterPro; IPR000815; Hg_reductase.  
InterPro; IPR006258; Lipamide dh.  
InterPro; IPR001100; Pyr_redox.  
InterPro; IPR004099; pyr_redox.dim.  
InterPro; IPR000103; Pyridine_redox_2.  
Pfam; PF000070; pyr_redox; 1.  
Pfam; PF02852; pyr_redox_dim; 1.  
PRINTS; PR00368; FADPNR.  
PRINTS; PR00945; HGDRASE.  
PRINTS; PR00411; PNDRDTASEI.  
PRINTS; PR00469; PNDRDTASEII.  
ProDom; PD000139; FAD_pyr_redox; 1.  
TIGRFAMs; TIGR01350; lipamide DH; 1.  
PROSITE; PS00076; PYRIDINE_REDOX_1; 1.  
Redox-active center; Glycylsis; Oxidoreductase; NAD; Flavoprotein;  
FAD; Complete proteome.  
NP BIND 34 42 FAD (ADP PART) (BY SIMILARITY).  
DISULFID 42 47 REDOX-ACTIVE (BY SIMILARITY).  
ACT_SITE 436 436 BY SIMILARITY.  
CONFLICT 64 64 K -> T (IN REF. 5).  
SEQUENCE 461 AA; 49166 MW; CAD4483E758FFAD CRC64;  
  
Query Match 7.7%; Score 82; DB 1; Length 461;  
Best Local Similarity 20.4%; Pred.No.25;  
Matches 46; Conservative 44; Mismatches 69; Indels 66; Gaps 11;  
  
QY 23 DKITEE----INKAIDDATIAAEIQSEITDPMKVPDHAKFE-----RH 61  
DB 219 NKFTQSGIRLTKA---SSAISAESQNQRVITNQVEEFVLVAIGQNTASIGLDN 275  
QY 62 VGIV-DPKGELAMENTEARGLKOMKQQG-----ANVKSEGVIKHAHLIGVHHDIVS 113  
DB 276 AGVIDRDGRVIPDETRTNVNPIYAIGDTGKKLLAHVASHGVIATAKNISGHHE---V 332  
QY 114 MEYDLAVKLGDLPHTTHVISDIQDFVALSLISEGNI--TWTSFEVRQ----- 161  
DB 333 MDYSAPFSVFIFTH-----BIAMVGLSLOEAQQNLPAKLTFPPPKAIKGKAVLGAS 384  
QY 162 --FAMVVNH-----IGGLSILDPIFGVLSVDLTAIFTQDTVREMT 199  
DB 385 DGFAIVSHETQQILGAIVIGHASSLIGEWTL----AIRNELT 425  
  
RESULT 49  
MVS3 SCHPO  
ID _MVS3 SCHPO STANDARD; PRT; 2104 AA.  
AC O14157; O42730; Q9UJ49;  
DT 16-OCT-'2001 (Rel. 40, Created)  
DT 16-OCT-'2001 (Rel. 40, Last sequence update)  
DT 15-MAR-'2004 (Rel. 43, Last annotation update)  
DE Myosin type II heavy chain 2.
```

GN MYO3 OR MYP2 OR SPAC4A8.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98119390; PubMed=9459302;
 RA Motegi F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
 RT "Identification of Myo3, a second type-II myosin heavy chain in the
 RL fission yeast Schizosaccharomyces pombe."; *FEBs Lett.* 420:161-166(1997).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98062346; PubMed=9398685;
 RA Bezanilla M., Forsburg S.L., Pollard T.D.;
 RT "Identification of a second, myosin-II in Schizosaccharomyces pombe:
 RL Myo3p is conditionally required for cytokinesis."; *Mol. Biol. Cell* 8:2693-2705(1997).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds R., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Taylor K., Taylor R.G., Tvey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Borzym K., Langer I., Beck A., Lehach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jinenz J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RL "The genome sequence of Schizosaccharomyces pombe."; *Nature* 415:871-880(2002).
 RN (4)
 RP SEQUENCE OF 1871-2078 FROM N.A.
 RX STRAIN=968 b30;
 RX MEDLINE=20223868; PubMed=10759889;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RL fission yeast cells by the use of a GFP-fusion genomic DNA library."; *Genes Cells* 5:169-190(2000).
 CC -!- FUNCTION: Stabilizes the F-actin cables forming the F-actin ring
 CC that surrounds the nucleus during interphase. May work in
 CC conjunction with myo2.
 CC -!- SUBUNIT: Binds to cdc4 and rlc1.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB007633; BAA24579.1; -;
 DR EMBL: AF029788; AAC04615.1; -;
 DR EMBL: Z38782; CAB11475.1; -;
 DR EMBL: AB027812; BAA87116.1; -;
 DR FIR: T38774; T38774.
 DR HSP: P08799; 1MND.
 DR GeneDB SPombe; SPAC4A8.05c; -;
 DR InterPro: IPR001609; myosin head.
 DR InterPro: IPR004009; myosin_N.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF02736; myosin_N; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00242; MYSC; 1.
 KW Myosin; Coiled coil; ATP-binding; Actin-binding.
 FT DOMAIN 1 828 MYOSIN HEAD-LIKE.
 FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
 FT NP_BIND 646 660 ACTIN-BINDING (POTENTIAL).
 FT NP_BIND 182 189 ATP (POTENTIAL).
 FT CONFLICT 1193 1193 D -> G (IN REF. 2).
 FT CONFLICT 1304 1304 E -> G (IN REF. 2).
 FT CONFLICT 1344 1344 E -> K (IN REF. 2).
 FT CONFLICT 1420 1420 G -> D (IN REF. 2).
 FT CONFLICT 1420 1420 G -> D (IN REF. 2).
 SQ SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;
 Query Match 7.7%; Score 82; DB 1; Length 2104;
 Best Local Similarity 24.8%; Pred. No. 1.5e+02;
 Matches 36; Conservative 25; Mismatches 54; Indels 30; Gaps 6;
 QY 24 KITEINKAIDATAAIEQSITIDPMKYPDHADKFERHGVDFKGLAMEN----- 75
 DB 1864 EVSEQNKKFDSSARPP-----DEMEKEVITLTKSNIAQKDDLLSLVE 1908
 QY 76 -TEARGLKQKQGDVANKYEGEVKVAHLIGVHDDIVMEYDIAKGLDLPHTHVID 134
 DB 1909 RIKQIEFALTKQDSNNHRENL-QLHRLGV---LQEKKKDLKLFLDLDTYPST 1964
 QY 135 IQDFVVALSLISD-EGNITWTSFE 158
 DB 1965 SKD-VRMLQKQISDLEASFAASDIE 1988
 RESULT 50
 CH60 PANAN
 ID CH60 PANAN STANDARD; PRT; 541 AA.
 AC O66218;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
 GN GROEL OR GROEL OR MOPA.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX STRAIN=JCM 6986;
 RA Harada H., Ishikawa H.;
 RT "Phylogenetical relationship based on groE genes among phenotypically
 RL related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
 RL species."; *J. Gen. Appl. Microbiol.* 43:355-361(1997).
 CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
 CC proper assembly of unfolded polypeptides generated under stress
 CC conditions (By similarity).
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 CC 7 subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB008151; BAA29235.1; --
 DR HSSP; P06139; 1DK7.
 DR HAMAP; MF_00600; --; 1.
 DR InterPro; IPR001844; Chaprinin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 KW Chapterone; ATP-binding.
 FT NON_TER 541
 SQ SEQUENCE 541 AA; 56817 MW; 7111AA03A703B543 CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 541;
 Best Local Similarity 18.1%; Pred. No. 34;
 Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDA-IAAIEQSEITDPMKVPDHPADKFERHVGIVDFKGL 71

DB 107 VAAGNPM-----DLKRGIDQAVAAVEE---LKLSPVCSDSKAAIAQVGTISANSDE 156

QY 72 AMENIEARGLKQMKQGDANVKEGIVKAHLIGVHD--IVSNEYDLY----- 120

DB 157 TVGTLLAQAMEKV-----GKEGVITVEGTGLQDELVDVEGMQDFRGYSFYFINK 207

QY 121 -----KLGLHPHTTHVSDIQDFVVALSLEISDEGNITWTSFV 159

DB 208 PETGAVEETPPILLADKKISNIREMLPVLEAVAKAGKPLLIADVEGEALATL----- 262

QY 160 ROPANYVNHIGGL-----SILDPIFGVLSVLTAFQDITVRKEMTKVLAFAFKRELEK 212

DB 263 -----VNTMTRGIVKVAAVKAPGFG---DRRKAMLDIAILTGGTVISEIGMELEK 311

RESULT 51

CH60_HAEDU

ID CH60_HAEDU STANDARD; PRT; 547 AA.

AC P31294;

DT 01-JUL-1993 (Rel. 26, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).

GN GROEL OR GROEL OR MOPA OR HD1784.

OS Haemophilus ducreyi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=730;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=93014108; PubMed=1356926;

RA Parsons L.M., Waring A.L., Shayegani M.;

RT "Molecular analysis of the Haemophilus ducreyi groE heat shock

operon";

RL Infect. Immun. 60:4111-4118(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=35000HP / ATCC 700724;

RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,

RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;

RT "The complete genome sequence of Haemophilus ducreyi";

RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases

CC -!- FUNCTION: Prevents misfolding and promotes the refolding and

proper assembly of unfolded polypeptides generated under stress

conditions (By similarity).

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

7 subunits (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; M91030; AAA24961.1; --
 DR EMBL; AE017155; AAB96536.1; --
 DR PIR; B49203; B49203.
 DR HSSP; P06139; 1GRL.
 DR HAMAP; MF_00600; --; 1.
 DR InterPro; IPR001844; Chaprinin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60_TCP1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 KW Chapterone; ATP-binding; Complete proteome.
 FT CONFLICT 37 N -> K (IN REF. 1).
 FT CONFLICT 37 A -> V (IN REF. 2).
 FT CONFLICT 509 509 A -> V (IN REF. 2).
 SQ SEQUENCE 547 AA; 57798 MW; B9DF933AD998B5C8 CRC64;

Query Match 7.6%; Score 81.5; DB 1; Length 547;
 Best Local Similarity 18.2%; Pred. No. 34;
 Matches 43; Conservative 44; Mismatches 82; Indels 67; Gaps 9;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAIEQSEITDPMKVPDHPADKFERHVGIVDFKGLA 72

DB 107 VAAGNPM-----DLKRGIDQAVAAV--VEELKASKCETSKEIEQVGTISANSDET 157

QY 73 MNIEARGLKQMKQGDANVKEGIVKAHLIGVHD--DIV-SNEYDLY----- 120

DB 158 VGKLLAQAMEKV-----GKEGVITVEDGTGLDLDVVEGMQDFRGYSFYFINKP 208

QY 121 -----KLGLHPHTTHVSDIQDFVVALSLEISDEGNITWTSFVR 160

DB 209 EAGTVELDNPIIIVDKKISNIREILPVLEAVAKAGKPLLIADVEGEALATL----- 262

QY 161 QFANYVNHIGGL-----SILDPIFGVLSVLTAFQDITVRKEMTKVLAFAFKRELEK 212

DB 263 -----VNTMTRGIVKVAAVKAPGFG---DRRKAMLDIAILTAGTVISEIGMELEK 311

RESULT 52

CH61_VIEHA

ID CH61_VIEHA STANDARD; PRT; 548 AA.

AC Q83W18;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE 60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1).

GN GROEL OR GROEL1.

OS Vibrio Harveyi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=669;

RN [1]

RP SEQUENCE FROM N.A.

RA Kuchanny-Ardigo D., Lipinska B.;

RT "Cloning and characterization of the groE heat shock operon of the

marine bacterium Vibrio Harveyi";

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Prevents misfolding and promotes the refolding and

proper assembly of unfolded polypeptides generated under stress

conditions (By similarity).

CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of

7 subunits (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AY246431; AA088905.1; -;
 DR HAMAP; MF 00600; -; 1.
 DR InterPro; IPR001844; Chaperonin Cpn60.
 DR InterPro; IPR002423; Cpn60/TCF-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; cpn60_TCF1; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTC1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 KW Chaperone; ATP-binding.
 KW SEQUENCE 548 AA; 57549 MW; F37B834D7A60C45 CRC64;
 CC
 CC Query Match 7.6%; Score 81.5; DB 1; Length 548;
 CC Best Local Similarity 25.0%; Pred. No. 34;
 CC Matches 42; Conservative 18; Mismatches 71; Indels 37; Gaps 5;
 CC
 CC 6 LIAAVAFVAVSA-----DPIHYDKITEINKAIDDAIAAIEQSETIDPMKV 51
 CC 372 LAGGVAIVKGAATEVEMKEKKRVEBALHATRAAVEGVAGGVALIRAASKI-----V 427
 CC
 CC 52 PDHADKPERHVGIVDFKGLAMRNIEARGLQKMKRQGDA-----NVKGEGEVKAHLII 105
 CC 428 DLEGNEEQNGVI-----RVALRAEAPIRQITKNAGDEESVANNVKAGEGSGVGNAT 482
 CC
 CC 106 GVHDDIVSMEDLAYKGLDLPHTTHVSDIQDFVALSLISBEGNT 153
 CC 483 GEYGDMLDEM-----GILDPTKVTRSALQFAASVAGLMTITTEAMVT 522
 CC
 CC RESULT 53
 CC ATCS_SVNY3
 CC ID ATCS_SVNY3 STANDARD; PRT; 745 AA.
 CC AC P73241;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Cation-transporting ATPase pacs (EC 3.6.3.-).
 CC GN PACS OR SLL1920.
 CC OS Synechocystis sp. (strain PCC 6803).
 CC OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 CC ON NCBI_TaxID=11148;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=97061201; PubMed=8905231;
 CC RA Kaneko T., Sato S., Kori H., Tanaka A., Asamizu E., Nakamura Y.,
 CC RA Miyajima N., Hirose A., Sugita M., Sugita M., Sasaki T.,
 CC RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 CC RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 CC RA Yamada M., Yasuda M., Tabata S.;
 CC RT "Sequence analysis of the genome of the unicellular cyanobacterium
 CC RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 CC RT entire genome and assignment of potential protein-coding regions.";
 CC EL DNA Res 3:109-136(1996).
 CC CC -!- FUNCTION: MAY PLAY A ROLE IN THE OSMOTIC ADAPTATION (BY
 CC SIMILARITY).
 CC CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IB.
 CC CC -!- SIMILARITY: Contains 1 HMA domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90904; BAA17268.1; -;
 DR PIR; S75354; S75354.
 DR HSP; P04129; IAFU.
 DR InterPro; IPR006403; ATPase-IB1 Cu.
 DR InterPro; IPR006416; ATPase-IB hvy.
 DR InterPro; IPR001757; ATPase-E1-E2.
 DR InterPro; IPR001756; Cu ATPase.
 DR InterPro; IPR008250; E1-E2 ATPase reg.
 DR InterPro; IPR006121; HeavyMe_transpt.
 DR InterPro; IPR001802; HG scavenger.
 DR InterPro; IPR005834; Hydrolase.
 DR InterPro; IPR006191; Metal bind.
 DR Pfam; PF00122; E1-E2 ATPase; 1.
 DR Pfam; PF00403; HMA; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR PRINTS; PR00943; CUATPASE.
 DR PRINTS; PR00946; HGSCAVENGER.
 DR TIGRFAMS; TIGR01511; ATPase-IB1 Cu; 1.
 DR TIGRFAMS; TIGR01525; ATPase-IB hvy; 1.
 DR TIGRFAMS; TIGR01494; ATPase P-type; 2.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PROSITE; PS01047; HMA_1; 1.
 DR PROSITE; PS0846; HMA_2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Metal-binding; Complete proteome.
 KW DOMAIN 1 94
 FT TRANSMEM 95 115
 FT DOMAIN 116 125
 FT TRANSMEM 126 145
 FT DOMAIN 146 152
 FT TRANSMEM 153 173
 FT DOMAIN 174 193
 FT TRANSMEM 194 214
 FT DOMAIN 215 342
 FT TRANSMEM 343 365
 FT DOMAIN 366 372
 FT TRANSMEM 373 390
 FT DOMAIN 391 543
 FT TRANSMEM 544 564
 FT DOMAIN 565 687
 FT TRANSMEM 688 707
 FT DOMAIN 708 719
 FT TRANSMEM 720 738
 FT DOMAIN 739 745
 FT DOMAIN 745 745
 FT MOD_RES 428 428
 FT METAL 14 14
 FT METAL 17 17
 FT METAL 633 633
 FT METAL 637 637
 SQ SEQUENCE 745 AA; 79950 MW; 40831F66621735FB CRC64;
 CC
 CC Query Match 7.6%; Score 81.5; DB 1; Length 745;
 CC Best Local Similarity 25.2%; Pred. No. 49;
 CC Matches 30; Conservative 26; Mismatches 46; Indels 17; Gaps 5;
 CC
 CC 77 EARGLKQMKRGDANVKGEGIVKAHLI-----IGVHDDIVSMEDLAYKGLDLPHTTHV 132
 CC 614 KAQQVEQLQKGNIVAMVGDGNDAPALAQADVGI---AIGTGDVAIASDI---TLIA 667
 CC
 CC 133 SDIQDFVALSLISBEGNTMTSFEVRQFANVNHIG---GLSILDPITFGLSDVLTA 188
 CC 668 GDLQGLITAKLSRATNGNTRQNLF----FAFYIVIGIPVAGLFPFLGLLNPLIA 722

RESULT 55
GID_AQUAE
ID GID
AC 0669

```
RESULT 56
ID UNR_HUMAN STANDARD; PRT; 798 AA.
AC O75534; O94961; O9Y294;
DT 30-MAY-2000 (Rel. 39, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE UNR protein.
GN UNR OR D15155E OR KIA0885.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10049485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364 (1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Pancreatic;
RA Song H., Peng Y., Zhou J., Huang Q., Dai M., Mao Y., Xu X.,
RA Luo M., Hu R., Chen J.;
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Cervix;
EX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altshuler S.F., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [4]
RP SEQUENCE OF 1-498 FROM N.A. (ISOFORM SHORT).
RC TISSUE=Brain;
RA Xu W., Gibbs R.A.;
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP INTERACTON WITH UNRIP.
RX MEDLINE=99158740; PubMed=10049359;
RA Hunt S.L., Huan J.J., Totty N., Jackson R.J.;
RT "unr, a cellular cytoplasmic RNA-binding protein with five cold-shock
RT domains, is required for internal initiation of translation of human
RT rhinovirus RNA.";
RL Genes Dev. 13:437-448 (1999).
RN [6]
RP -!- FUNCTION: RNA-binding protein. Required for internal initiation of
RP translation of human rhinovirus RNA.
RN [7]
RP -!- SUBUNIT: Interacts with UNRIP.
RN [8]
RP -!- SUBCELLULAR LOCATION: Cytoplasmic.
RN [9]
RP -!- ALTERNATIVE PRODUCTS:
```

```
DE Bacterioferritin (BFR) (Cytochrome B-1) (Cytochrome B-557).
GN BFR.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RA Noorani S.M., Lindahl L., Zengel J.M.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May perform analogous functions in iron detoxification
CC and storage to that of animal ferritins (By similarity).
CC -!- COPACTOR: Binds 1 heme B (iron-protoporphyrin IX) group per dimer
CC (By similarity).
CC -!- SUBUNIT: Oligomer of 24 identical subunits (By similarity).
CC -!- MISCELLANEOUS: The di-iron binding site functions as active site
CC where iron ions are oxidized from iron(II) to iron(III) before
CC they are stored (By similarity).
CC -!- SIMILARITY: Belongs to the bacterioferritin family.
CC -!- SIMILARITY: Contains 1 ferritin-like diiron domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC ENBL; AF058451; AAC14293.1; -.
CC HSPSP; P11056; IBCF.
CC InterPro; IPR002024; Bacterioferritin.
CC InterPro; IPR008331; Ferritin Dps.
CC InterPro; IPR009040; Ferritin_like.
CC Pfam; PF00210; ferritin; 1.
CC PRINTS; PR00601; BACFERRITIN.
CC PRODOM; PD002269; Bacterioferritin; 1.
CC TIGRFAMS; TIGR00754; bfr; 1.
CC PROSITE; PS00549; BACTERIOFERRITIN; 1.
CC PROSITE; PS00905; FERRITIN LIKE; 1.
CC Iron storage; Heme; Metal-binding; Iron.
FT DOMAIN 1 145 FERRITIN-LIKE DIIRON.
FT METAL 18 18
FT METAL 51 51 IRON 1 (BY SIMILARITY).
FT METAL 51 51 IRON 1 (BY SIMILARITY).
FT METAL 51 51 IRON 2 (BY SIMILARITY).
FT METAL 52 52 IRON 2 (BY SIMILARITY).
FT METAL 54 54 IRON 1 (BY SIMILARITY).
FT METAL 54 54 IRON 2 (BY SIMILARITY).
FT METAL 94 94 IRON 1 (BY SIMILARITY).
FT METAL 127 127 IRON 2 (BY SIMILARITY).
FT METAL 127 127 IRON 2 (BY SIMILARITY).
FT METAL 130 130 IRON 2 (BY SIMILARITY).
SQ SEQUENCE 159 AA; 18533 MW; 831A865EA98C382D CRC64;
Query Match 7.5%; Score 80.5; DB 1; Length 159;
Best Local Similarity 24.6%; Pred. No. 9.7;
Matches 41; Conservative 19; Mismatches 46; Indels 61; Gaps 8;
QY 24 KITTEINKAIDDAIAAEQ-----SETIDPMKVPDHDADKFERHV 62
DB 6 KITIAHLNKLGLNELVALNQVFLHARFKNWGLRLNDKEYHESIDEMK---HADRYIERI 62
QY 63 ----GIYDFKGEIARNT-----EARGLKMKRGDANVKGEIGVKAHL 104
DB 63 LFLEGIPNLQ-DLQKLNIGEDIEMLRSDLALELAGAKNL-REGIAYADSIHDYVSRDLM 120
QY 105 IGV-----HDDIVSMYDLYKGLDLPHTTHVISDQDFVALSLE 145
DB 121 IDILADBEHIDWLETDLIARLG-----IQNYAQAIL 156
RESULT 58
VATE_SULTO
ID_VATE_SULTO STANDARD; PRT; 194 AA.
AC Q971B8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V-type ATP synthase subunit E (EC 3.6.3.14) (V-type ATPase subunit E).
GN ATPB OR ST1435.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=JCM 10545 / 7;
EX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7."
RL DNA Res. 8:123-140(2001).
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SIMILARITY: Belongs to the V-ATPase E subunit family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC ENBL; AF000986; BAB66502.1; -.
CC HAMAP; MF_00311; -.
CC InterPro; IPR002842; ATPsynth Esub.
CC Pfam; PF01991; V-ATP-synt E; 1.
CC Hydrolase; ATP synthesis; Hydrogen ion transport;
KW Complete proteome.
SQ SEQUENCE 194 AA; 22634 MW; 20F46AF95783449A CRC64;
Query Match 7.5%; Score 80.5; DB 1; Length 194;
Best Local Similarity 20.6%; Pred. No. 12;
Matches 41; Conservative 39; Mismatches 68; Indels 51; Gaps 9;
QY 22 YDKITEINKAIDDAIAAEQSEITIDPMKVPDHDADKFERHVGVDFKGEIARNTAEGL 81
DB 26 FKILSEMNQIDEAYAEVYREYS---AKITDLVNKNNDRI-----RGEIARMEIENKRL 77
QY 82 KMKRGQDANVKGEIGVKAHLIGVHDDIVSMYDLYKGLDLPHTTHVISDQDFVVA 141
DB 78 --ISKEMDYIENVKKNKSLYEFVKTD-----NYKKG-----LESI----- 113
QY 142 ISLEISDEGNTMTSFEVRQFANVNH-----IGGLSILDPFGVLSLSD----- 184
DB 114 ISREVSDESGIIVCSPSDQKSIDIKKKIKSVWDEKIVGGIKYIYDPKSLSKDFTLE 173
QY 185 -VLTAFIDFTVRKEMTKVL 202
DB 174 TILNQVFD-IRDKIAQIL 191
RESULT 59
CH60_ENTAG
ID_CH60_ENTAG STANDARD; PRT; 539 AA.
AC O66200;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOPA.
OS Enterobacter agglomerans (Pantoea agglomerans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 1236;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008142; BAA25217.1; -
CC HSSP; P06139; 1JON.
CC HAMAP; MF 00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC NON_TER 539 539
CC SEQUENCE 539 AA; 56373 MW; 6583295F73299C3B CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 539;
Best Local Similarity 18.1%; Pred. No. 40;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPHYDKITEINKAIDDAIA-AIEQSEITDPMKVPDHDADKPERHVGIVDPKGE 71
DB 107 VAGNPNM-----DLKRGIDKAVAAVEE---LKALSVPCSDSKAIAQVGTISANSDE 156

QY 72 AMRNIARGLQMKRQGDANVKGEEGVKAHLIGVHDD---IVSMYEDLAY----- 120
DB 157 TVGKLIAEAMDKV-----GKEGIVTVEDGGLDELDVVEGQDFRGYLSPYFINK 207

QY 121 -----KLGDLHPTHTVSDIQDFVVALSLEISDEGNITWTSFEV 159
DB 208 PETGAVELESFPILLADKKISNIREMLFVLEAVAKAGKPLVIAEDVEGEALATL----- 262

QY 160 RQFANVNVHIGGL-----SILDPFGVLSVDVLTALFQDVTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VNTMREGIVKVAAPKAPGFG---DRKAMQLQDIATLTGGTVISEEIGMELEK 311

RESULT 60
CH60_ENTAS
ID CH60_ENTAS STANDARD; PRT; 539 AA.
AC O66190;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).

GN GROEL OR GROEL OR MOPA.
OS Enterobacter asburiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=61645;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 6051;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species.";
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB008137; BAA25207.1; -
CC HSSP; P06139; 1JON.
CC HAMAP; MF 00600; -; 1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC NON_TER 539 539
CC SEQUENCE 539 AA; 56399 MW; 8A8A5CAF9ADBDAFB CRC64;

Query Match 7.5%; Score 80.5; DB 1; Length 539;
Best Local Similarity 18.1%; Pred. No. 40;
Matches 43; Conservative 43; Mismatches 82; Indels 69; Gaps 9;

QY 13 VAVSADPHYDKITEINKAIDDAIA-AIEQSEITDPMKVPDHDADKPERHVGIVDPKGE 71
DB 107 VAGNPNM-----DLKRGIDKAVAAVEE---LKALSVPCSDSKAIAQVGTISANSDE 156

QY 72 AMRNIARGLQMKRQGDANVKGEEGVKAHLIGVHDD---IVSMYEDLAY----- 120
DB 157 TVGKLIAEAMDKV-----GKEGIVTVEDGGLDELDVVEGQDFRGYLSPYFINK 207

QY 121 -----KLGDLHPTHTVSDIQDFVVALSLEISDEGNITWTSFEV 159
DB 208 PETGAVELESFPILLADKKISNIREMLFVLEAVAKAGKPLVIAEDVEGEALATL----- 262

QY 160 RQFANVNVHIGGL-----SILDPFGVLSVDVLTALFQDVTVRKEMTKVLAPAFKRELEK 212
DB 263 -----VNTMREGIVKVAAPKAPGFG---DRKAMQLQDIATLTGGTVISEEIGMELEK 311

RESULT 61
CH60_ENTAS
ID CH60_ENTAS STANDARD; PRT; 540 AA.
AC O66198;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN GROEL OR GROEL OR MOPA.

Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC -1- TISSUE SPECIFICITY: Pharyngeal muscle.

CC -1- MISCELLANEOUS: This protein binds two calcium ions (Potential).

CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; AB079299; BAB84566.1; -

DR EMBL; Z48585; CAA8482.1; -

DR PIR; T27963; T27963.

DR HSP; P02588; IPON.

DR WormPep; ZK673.7; CE01719.

DR InterPro; IPR003299; Calflagin.

DR InterPro; IPR002048; EF-hand.

DR Pfam; PF00036; ehand; 4.

DR PRINTS; PR01362; CALFLAGIN.

DR ProDom; PD000012; EF-hand; 2.

DR SMART; SM00054; Eph; 4.

DR PROSITE; PS00019; EF_HAND; 2.

KW Calcium-binding; Repeat.

FT DOMAIN 28 39 ANCESTRAL CALCIUM SITE 1 (POTENTIAL).

FT CA BIND 64 75 EF-HAND 2 (POTENTIAL).

FT DOMAIN 105 116 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).

FT CA BIND 141 152 EF-HAND 4 (POTENTIAL).

SQ SEQUENCE 160 AA; 18227 MW; 18552E2D8D5D58CE CRC64;

Query Match 7.5%; Score 80; DB 1; Length 160;

Best Local Similarity 23.1%; Pred. No. 11;

Matches 27; Conservative 28; Mismatches 42; Indels 20; Gaps 3;

QY 53 DHADKFERHVGIVDFKGLAMR-----NIEARGLKMKRQGDANVKGE--- 95

Db 15 DQIEQFKYVMPDKGKGVIRATQVQILRTWGQAFERDLKLIKEFDAGSGEIEPE 74

QY 96 --EGIVKAHLIGVHDIVSMYDLYKGLDHPHTHV-ISDIQDFVVALSLEISDE 149

Db 75 EFAAMVANFYVNNENDEGLEFEELREAFRLYDKGNGYINVSILRLDILRALDNNVSEE 131

RESULT 66

SYTB_METKA

AC Q8TX56; STANDARD; PRT; 590 AA.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)

DE (Phenylalanine--tRNA ligase beta chain) (PheRS).

GN PHE1 OR MK0820.

OS Methanopyrus kandleri.

OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;

OC Methanopyrus.

OX NCBI_TaxID=2320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;

RA MEDLINE=21927647; PubMed=11930014;

RA Slesarev A.I., Mezhevaya K.V., Belova G.I., Aravind L., Shcherbinina O.V., Shakhova V.V., Tatusov R.L., Wolf Y.I., Stetter K.O., Ntale D.A., Rogozin I.B., Kozaykin S.A., Malykh A.G., Koonin E.V., Kozaykin S.A.;

RT "the complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens";

RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).

CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP + diphosphate + L-phenylalanyl-tRNA(Phe).

CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By

similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain family. Subfamily 2.

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; AE010373; AAM02033.1; -

DR HAMAP; MF 00284; -; 1.

DR InterPro; IPR005146; B3_4.

DR InterPro; IPR005147; B5.

DR InterPro; IPR004531; PheT_arch.

DR Pfam; PF03483; B3_4; 1.

DR Pfam; PF03484; B5; 1.

DR TIGRFAMs; TIGR00471; pheT_arch; 1.

KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW Complete proteome.

SQ SEQUENCE 590 AA; 67511 MW; FE1D644780F256BE CRC64;

Query Match 7.5%; Score 80; DB 1; Length 590;

Best Local Similarity 22.6%; Pred. No. 49;

Matches 55; Conservative 28; Mismatches 74; Indels 86; Gaps 12;

QY 20 IYDKITEEINKADDAATAEQSETIDPMKVPD--HADKFERHVGIVDFKGLAMRNI 77

Db 6 IYDKIV-----KILGREVSFEELAHNLIPLMGSDVERIDEREMVETETFFNRPDLISVE 61

QY 78 --ARGLKQM-----KRGDANVKGEIVKA-----HL 103

Db 62 GVARALKGFLGTCIPYNNRRSDVEARVESVLDARPCLAVALVGVGVEPDERDLHL 121

QY 104 L-----IGVHDDIVSMYDLYKGLDHPHTHV-ISDIQDFVVALSLE 145

Db 122 MEFOHHLHWVGRDKKAAIGH--DFEAVEPLRYFLAD--PNDR-----SWAFE 168

QY 146 ISDEGNITMTSFEV-----RQFANVNVNHIGLSILDPIFGVLSVDVLTAFQDTVRKEM 198

Db 169 PLDHPGEMTPAEVLRLRHKRGQYAHVLS--DGAPILADEBEGVIS-----FPPVINSE 220

QY 199 TKV 201

Db 221 TRV 223

RESULT 67

UNR_RAT

ID UNR_RAT STANDARD; PRT; 798 AA.

AC P18395;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE UNR protein.

DE UNR.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA MEDLINE=90370473; PubMed=2204029;

RA Jeffers M., Pacucci R., Pellicer A.;

RT "Characterization of unr; a gene closely linked to N-ras.";

RL Nucleic Acids Res. 18:4891-4899 (1990).

CC -1- FUNCTION: RNA-binding protein (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to the cold-shock domain (CSD) family.

Db	208	ETGAVELESFILLADKKISNIREMLPVLEAVAKAGKPLIIIAEDVEGEMATL-----	261		
Qy	161	QFANVNHIGGL-----SILDPFGVLSVLTAFQDTRKEMTKVLAPAFKRELEK 212			
Db	262	-----VNTMRGIVKVAAPGFG---DRRKAMLODIATLTGGTVISEEIGMELEK 310			
RESULT 72					
LON_CHLTR					
ID	LON_CHLTR	STANDARD; PRT; 819 AA.			
AC	084348;				
DT	30-MAY-2000	(Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	ATP-dependent protease La	(EC 3.4.21.53).			
GN	LON OR CT344.				
OS	Chlamydia trachomatis.				
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=813;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=D/UM-3/Cx;				
RX	MEDLINE=99000809; PubMed=9784136;				
RA	Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;				
RT	"Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";				
RL	Science 282:754-759(1998).				
CC	-!- FUNCTION: Degrades short-lived regulatory and abnormal proteins in presence of ATP. Hydrolyzes two ATPs for each peptide bond cleaved in the protein substrate (By similarity).				
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin, casein and denaturated serum albumin, in presence of ATP.				
CC	-!- SUBUNIT: Homotetramer (By similarity).				
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-!- SIMILARITY: Belongs to peptidase family S16.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL; AE001307; AAC67939.1; --				
DR	PIR; C71527; C71527.				
DR	MEROPS; S16.002;				
DR	InterPro; IPR003593; AAA_ATPase.				
DR	InterPro; IPR003959; AAA_ATPase_centr.				
DR	InterPro; IPR001270; Chaprinin clpA/B.				
DR	InterPro; IPR008269; Pept_S16_C.				
DR	InterPro; IPR004815; Pept_S16_N.				
DR	InterPro; IPR003111; Pept_S16_N.				
DR	InterPro; IPR008268; Peptid_S16_AS.				
DR	InterPro; IPR001984; Peptidase_S16.				
DR	Pfam; PF00004; AAA; 1.				
DR	Pfam; PF02190; LON; 1.				
DR	Pfam; PF05362; Lon_C; 1.				
DR	PRINTS; PR00300; CLPPTOTASEA.				
DR	PRINTS; PR00830; ENDOLAPTASE.				
DR	SMART; SM00382; AAA; 1.				
DR	SMART; SM00464; LON; 1.				
DR	TIGRFAMS; TIGR00763; Lon; 1.				
DR	PROSITE; PS01046; LON_SER; 1.				
KW	Hydrolase; Serine protease; ATP-binding; Complete proteome.				
FT	NP BIND 392 399				
FT	ACT_SITE 724 724				
FT	BY SIMILARITY.				
SQ	SEQUENCE 819 AA; 91965 MW; FE47802A5441557C CRC64;				
Query Match	7.4%;	Score 79.5;	DB 1; Length 819;		
Best Local Similarity	21.1%;	Pred. No. 79;			
Matches	51; Conservative	35; Mismatches	73; Indels	83; Gaps	10;
Qy	16	SADPHYDKITEENKAIDDAIAAEQSETIDPMKVPDH-----	54		
Db	6	NTDSQNDPNASEVEKLIDESAEE-----KVDDHTPPSELFILPLNKRFFPGMA 57			
Qy	55	-----ADKFERHVGIVDFKGLAMENIARGLKQMKROCDANVKGEEG 97			
Db	58	APLIEAGPHYEVLTLLAKSSQKHGLVLTAKDA--NTLKVGFNQLHRVG-----	106		
Qy	98	IVKAHLIGVHDDIVSMYEDLAYKLGDLHPTTHVISDIQDFVVALSLEISDEGNITWTSF 157			
Db	107	-VSARIL-----RIMPIEGSAQVLLSTEDIRIVKPIQDKYLKAKVSYHKE-NKELTE- 158			
Qy	158	EVROFA-NVNVHIGGLSLDPIF-----GVLSVLTAFQDTRKEMTK 200			
Db	159	ELKAYSISIVSIIDLLKLNPLFKEELQIFLCHSDFTFPGKLADFSVALTTAT-REELQE 217			
Qy	201	VL 202			
Db	218	IL 219			
RESULT 73					
CTN2_CHICK					
ID	CTN2_CHICK	STANDARD; PRT; 906 AA.			
AC	P30937;				
DT	01-JUL-1993	(Rel. 26, Created)			
DT	01-JUL-1993	(Rel. 26, Last sequence update)			
DT	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	Alpha-2 catenin (Alpha N-catenin) (Neural alpha-catenin).				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Embryonic brain;				
RX	MEDLINE=92346716; PubMed=1638632;				
RA	Hirano S., Kimoto N., Shimoyama Y., Hirohashi S., Takeichi M.;				
RT	"Identification of a neural alpha-catenin as a key regulator of cadherin function and multicellular organization.";				
RL	Cell 70:293-301(1992).				
CC	-!- FUNCTION: ASSOCIATES WITH THE CYTOPLASMIC DOMAIN OF A VARIETY OF CADHERINS. THE ASSOCIATION OF CATENINS TO CADHERINS PRODUCES A COMPLEX WHICH IS LINKED TO THE ACTIN FILAMENT NETWORK, AND WHICH SEEMS TO BE OF PRIMARY IMPORTANCE FOR CADHERINS CELL-ADHERION PROPERTIES. ALPHA N-CATENIN IS CRUCIAL NOT ONLY FOR CADHERIN FUNCTION BUT ALSO FOR ORGANIZATION OF MULTICELLULAR STRUCTURES.				
CC	-!- TISSUE SPECIFICITY: MAINLY IN THE NERVOUS SYSTEM.				
CC	-!- SIMILARITY: STRONG, TO VINCULINS AND TO OTHER ALPHA-CATENINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL; D11090; BAA01863.1; --				
DR	PIR; A43000; A43000.				
DR	HSSP; P26231; LDOW.				
DR	InterPro; IPR001033; Alpha catenin.				
DR	InterPro; IPR006077; Vinculin catenin.				
DR	InterPro; IPR000633; Vinculin_2.				
DR	Pfam; PF01044; Vinculin; 1.				
DR	PRINTS; PR00805; ALPHACATENIN.				
DR	PRINTS; PR00806; VINCULIN.				
DR	PROSITE; PS00663; VINCULIN.1; 1.				
KW	Cytoskeleton; Structural protein; Cell adhesion.				
SQ	SEQUENCE 906 AA; 100690 MW; 7714D04918572DB7 CRC64;				


```

Query Match      7.4%; Score 79.5; DB 1; Length 906;
Best Local Similarity 19.4%; Pred.No.89;
Matches 48; Conservative 47; Mismatches 82; Indels 71; Gaps 10;

QY    18 DPH--YKITE-----EINKAIDDAIA-----AIEOSETIDPMKVDDHAD 56
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    361 DELNIALDKMTKTDRLRQLRKAVMDHISDSPLETNVPLLVLEAAKSNEKEVKYQA 420
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY    57 KPERHVG-IVDFKGELANRNTEARGLKQM-K-----RQ 87
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    421 VPFEHANKLVEVANLACSISSNEEGVLVMNAATQIDSLCPQVINAAULTAARPOSKVAQ 480
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY    88 GPANYKBEGEVGAHLITGVHDHDIVSMEYDLAYKLGLDLPHTTHVISDIODFVALSLEIS 147
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    481 DNNDVFQDQWEKVRVLTAEAVDDITSVDDELFSV-----ENHILEDNVKCIVA--LQE 532
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY    148 DEGNITMTSFPEVROPANVNHICGLSIDLDIFGVLSADVLTAFQTVRKEMTKVLA---- 203
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    533 DVDUDLRTAGAIRGRARVIHI-----INAMENYETGVYTEKV-LEATKLLSETVM 583
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

QY    204 PAKFRELE 211
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    584 PRFAEQVE 591
       |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

```

RESULT 74	ARGR_CLOTE	STANDARD;	PRT;	153 AA.
ID	ARGR_CLOTE			
AC	Q89AH3;			
DT	15-MAR-2004 (Rel. 43, Created)			
DT	15-MAR-2004 (Rel. 43, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Arginine repressor.			
GN	ARGR OR C1C01572.			
OS	Clostridium tetani.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1513;			
[1]	RN			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Massachusetts / E88;			
RX	MEDLINE=22457253; PubMed=12552129;			
RA	Bruggemann H., Baeumer S., Fricke W.F., Wierse A., Liesegang H.,			
RA	Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,			
RA	Gottschalk G.;			
RT	"The genome sequence of Clostridium tetani, the causative agent of			
RT	tetanus disease."			
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).			
CC	FUNCTION: Regulates arginine biosynthesis genes (By similarity).			
CC	PATHWAY: Arginine biosynthesis; regulation.			
CC	SUBCELLULAR LOCATION: Cytoplasmic (Probable).			
CC	SIMILARITY: Belongs to the argR family.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcements			
CC	or send an email to license@sib-sib.ch).			
CC				
CC	EMBL; AE015941; AAC361119.1; -.			
DR	HAMAP; MF_00173; -; 1.			
DR	InterPro; IPR001669; Arg repressor.			
DR	InterPro; IPR009058; Wing hlx DNA_bnd.			
DR	Pfam; PF01316; Arg repressor; 1			
DR	Pfam; PF02863; Arg_repressor; C; 1.			
DR	PRINTS; PR01467; ARGREPRESSOR.			
DR	Transcription regulation; DNA-binding; Trans-acting factor; Repressor;			
KW	Arginine biosynthesis. Complete proteome.			
SK	SEQUENCE 153 AA; 17120 MW; 27694F52058CF26 CRC64;			

```

Query Match          7.4%; Score 79; DB 1; Length 153;
Best Local Similarity 24.7%; Pred. No. 12;
Matches 42; Conservative 30; Mismatches 60; Indels 38; Gaps 9;

QY 49 MKVPDHADKFERHVGIVDFKG-----ELAMRNIE-----ARGLKMKROGDANVK 93
DB 1 MKVTRH---EKILELIERKDIOEBELAEELRKSGEITQATVSRDIKELKL---IKVL 53
QY 94 GEEGIVKAHLILGVHDDIVSMEYDLAKVIGDLHPHTHV-ISDIOQFVVALSLEISDEGNI 152
DB 54 GEQGYK-----YAAIVKNENLSDKLANIPSHSVSVSENNINNVVVKTL---SGSNA 104
QY 153 TMTSFYVRQFANVNVNHIGGISILDPIFGVLSDLVLTAFQDPTVRKEMTKVL 202
DB 105 AAAEAIIDSNFKETAGTAG---DNTIFIMARTSEQAFE-IVKKMKRVI 148

RESULT 75
ID KPYK_THELI
AC Q5c301; STANDARD; PRT; 220 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pyruvate kinase (EC 2.7.1.40) (PK)(Fragment).

```

```

RESULT 75
ID_KPYK THELI STANDARD; PRT; 220 AA.
AC Q56301;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Pyruvate kinase (EC 2.7.1.40) (PK) (Fragment).
DI PKI.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=2265;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DSM 5473;
RC MEDLINE=95370164; PubMed=7642512;
RA Kletzlin A., Mukund S., Kelley-Crouse T.L., Chan M.K., Rees D.C.,
RA Adams M.W.W.;
RT "Molecular characterization of the genes encoding the tungsten-
RT containing aldehyde ferredoxin oxidoreductase from Pyrococcus
RT furiosus and formaldehyde ferredoxin oxidoreductase from Thermococcus
RT litoralis.";
RL J. Bacteriol. 177:4817-4819 (1995).
CC -I- CATALYTIC ACTIVITY: ATP + Pyruvate = ADP + phosphoenolpyruvate.
CC -I- PATHWAY: Glycolysis, final step.
CC -I- SUBUNIT: Homotetramer (By similarity).
CC -I- SIMILARITY: Belongs to the pyruvate kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X83963; CAA58793.1; --
CC PIR; A57418; A57418.
CC HSPG; P11974; IPKN.
CC InterPro; IPR001697; Pyruvate_kinase.
CC Pfam; PF00224; PK; 1.
CC ProDom; PD001009; Pyruvate_kinase; 1.
CC PROSITE; PS00110; PYRUVATE_KINASE; PARTIAL.
CC Transferase; Kinase; Glycolysis; Magnesium.
FT NON TER 220
SQ SEQUENCE 220 AA; 24338 MW; A60B906BC742FCAA CRC64;

Query Match 7.4%; Score 79; DB 1; Length 220;
Best Local Similarity 23.0%; Pred. No; 19;
Matches 48; Conservative 36; Mismatches 71; Indels 54; Gaps 10;

Qy 19 PIHYDKI-----TTEINKAIDD--AIAAI-----EQSETIDMPKVPDHADKF 58
Db 8 PSHKTKIIATIGPASKOKESIKWIKAGMSVAINFSGHGLESHAKTIEVVR--DVAEKL 65

```

QY 59 ERHVGIVDFKGLAMRNIEARGIKQMKROD-----ANVKGEGIVKAHL-----LIGV 107
 Db 66 ERRVAILGDIPLGKVRGKIKGDSVTLRKDGKXVLTTRDIEGDETTIPVEFKDLPKLVSK 125
 QY 108 HDDIVSMEYDLAYKGLDHPHTTHVSDIQDFVVALSLEISDEG-----NITMTS 156
 Db 126 GDTIYLSGDIYMLREVEVRE-----NEVECVVNGGILFSGHKGINIPKANLPLEAITPRD 180
 QY 157 FEVQFANVNVH-----IGGLSILDPIDFVL 182
 Db 181 FEIIEFA--IEHGVDAIGLSFGVSGVDVL 207

RESULT 76
 RL4B_ARATH STANDARD; PRT; 406 AA.
 ID LE11 PYRAB
 AC Q9SF40;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L4-2 (L1).
 GN RPL4B OR AT3G09630 OR F118.22.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
 RA Faltmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Catolico L., Weissenbach J., Sauvin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Brille H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordisiek G.,
 RA Reichelt J., Schaefer M., Schoen O., Bagues M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy B.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfau G., Argirou A., Flores M., Liguori R., Vitale D.,
 RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Millscher J., Sellers P., Gill J.B., Feldblyum T.V.,
 RA Pruss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Igesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shimpou S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana."
 RL Nature 408:820-822(2000).
 CC -!- SIMILARITY: Belongs to the L4E family of ribosomal proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AC016661; AAF23293.1; -;
 DR InterPro; IPR002136; Ribosomal_L4/L1E.
 DR Pfam; PF00573; Ribosomal_L4; 1.

DR PROSITE; PS00939; RIBOSOMAL_L1E; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 406 AA; 44702 MW; A56AFA6CEAF291F6 CRC64;
 Query Match 7.4%; Score 79; DB 1; Length 406;
 Best Local Similarity 26.4%; Pred. No. 38;
 Matches 29; Conservative 24; Mismatches 41; Indels 16; Gaps 4;
 QY 1 MKKELLIAAFAVAVSADPI---HYDKITEINKAIDALAAIBQS-----ETIDPMKVPD 53
 Db 124 MKRHAIVSAIAATAVPAIWMARGHKIENVPMPVSDSAEAVEKTSAAIKVLKQIGAYD 183
 QY 54 HADKFERHVGIVDFKGLAMRNIEARGIKQMKROGDANVKGEG--IVKA 101
 Db 184 DAERAKNSIGIRPGKGM-----RNRVVISRKGPLVVTGEGSKIYKA 226

RESULT 77
 LE11 PYRAB
 ID LE11 PYRAB STANDARD; PRT; 486 AA.
 AC Q9UZ08;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 2-isopropylmalate synthase 1 (EC 2.3.3.13) (Alpha-isopropylmalate
 synthase 1) (Alpha-IPM synthetase 1).
 GN LEU1 OR PYRAB13490 OR PAB0890.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 CX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GES / Orsay;
 RX MEDLINE=22511545; PubMed=12622808;
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
 RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
 RA Van der Oost J., Weissenbach J., Zivanovic V., Forterre P.;
 RT "An integrated analysis of the genome of the hyperthermophilic
 archaeon Pyrococcus abyssi."
 RL Mol. Microbiol. 47:1495-1512(2003).
 CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
 CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-isopropylmalate) to form
 CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O =
 CC 2-hydroxy-2-isopropylsuccinate + CoA.
 CC -!- PATHWAY: Leucine biosynthesis; first step.
 CC -!- SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate
 CC synthase family. LeuA 1 subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ248287; CAB50254.1; -;
 DR PIR; A75045; A75045.
 DR HAMAP; MF_01025; -; 1.
 DR InterPro; IPR002034; AIPM/Hcit synth.
 DR InterPro; IPR000891; HMGL-like.
 DR InterPro; IPR005671; LeuA_bact_synth.
 DR Pfam; PF00682; HMGL-like; 1.
 DR TIGRfams; TIGR00973; leuA_bact; 1.
 DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
 DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
 DR Leucine biosynthesis; Transferrase; Complete proteome.
 SQ SEQUENCE 486 AA; 53768 MW; 721D5B4724CC49B2 CRC64;
 Query Match 7.4%; Score 79; DB 1; Length 486;
 Best Local Similarity 21.0%; Pred. No. 47;

```
Matches 58; Conservative 48; Mismatches 64; Indels 106; Gaps 18;
QY 12 FVAVSADPIHYD-----KITEINKAIDDAIAIEQSETIDPMKVPDPHADKFERHVGIVDF 67
Db 98 FIATS--PIHKYKLRKEPEVKKL---AVKAVE-----HATKYTED---VEF 137
QY 68 KGLAMRN-----IEARG-----LQMKRQGDANVKCEE 96
Db 138 SAEDATRSWDFLVEVEVAIDGATVINPDTGVATPEEYVELVRYLRNINISIKG-- 195
QY 97 GIVKAHLIGVHDDI--VSMEYDL--AYKLG--DLHPTTHVISD-----IQDFVVALSL-- 144
Db 196 ----VQISVHCDDDLGLAVANLSAISARAGADQVEVTNGIGERAGNAALEEIVVALDVR 251
QY 145 ----EISDEGNITWTSVEVQFANVNHIGLSILDPIFGVLSDLTA-----IPORTVKE 197
Db 252 DFVKVTDVNLK-----EIARTSKLVSHLTGIEV--PNKAIVGNAGFAHESGIHQDGVLKE 306
QY 198 MT--KVLP-----AFKRELEX 212
Db 307 RTTYEIIDPKLFGSGSKIVLGHSGRHAFRKLEE 342
RESULT 78
SYA_FUSNN
ID SYA_FUSNN STANDARD; PRT; 867 AA.
AC Q8RFJ8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (Alars).
GN ALAS OR FN0697.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Greckin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Lonsstein M., Kyripides N., Overbeek R.,
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5010580; AAL94893.1; ..
CC HAMAP; MF 00036; 1.
CC InterPro; IPR003156; DHHA1.
CC InterPro; IPR002318; tRNA-synt_2c.
CC InterPro; IPR006193; tRNA_synt_Ala.
CC Pfam; PF02272; DHHA1; 1.
CC Pfam; PF01411; tRNA-synt_2c; 1.
CC PRINTS; PR00980; TRNASYNTHALA.
CC TIGRFAMs; TIGR00344; alaS; 1.
CC PROSITE; PS00860; AA tRNA_LIGASE II ALA; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
```

```
SQ SEQUENCE 867 AA; 98147 MW; B941A2C030455F52 CRC64;
Query Match 7.4%; Score 79; DB 1; Length 867;
Best Local Similarity 23.0%; Pred. No. 92;
Matches 50; Conservative 37; Mismatches 80; Indels 50; Gaps 12;
QY 21 HYDKITE-----INKAIDDAIA-----AIEQSETIDPMKVPDPHADKFERHVGIVD 66
Db 592 HYEALTEQSLKIEDIVNKKIRIEGIVVVSHTSEAKLGAAML--FGDKYGVVVRWD 649
QY 67 ---FKGELA---NRNIEARGLKQKQGD--ANVKGEEGIV-----KAHLIGV 107
Db 650 VHGFSSTELCGGTHDNIGKIGLFPKITSEGGIAAGVRIEAKTGYGALVKEEADILKNI 709
QY 108 HDDIVSMYDLAYKGLDHPHTTHVISDIQDFVVALSLSEISDEGNITWTSVEVQFANVN 167
Db 710 EQKLKATNSNLVEK---VEKNLETKDKTEK-----ELEILKQ---KLALFETKAAISGME 758
QY 168 HIGLSILDPIFGVLS--DVLTAIFQDTVRKEMTKVL 202
Db 759 EIGGVKVLIAAFKDKSTEDLRTMI--DTIKDNEKAI 793
RESULT 79
LIPC_SOLTU
ID LIPC_SOLTU STANDARD; PRT; 326 AA.
AC P80471; Q24392;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Light-induced protein, chloroplast precursor (Chloroplastic drought-
DE induced stress protein CDSP-34).
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.
RP STRAIN=cv. Haig;
RX MEDLINE=95056847; PubMed=9839468;
RA Gillet B., Bayly A., Peltier G., Rey P.;
RT "Molecular characterization of CDSP 34, a chloroplastic protein
RT induced by water deficit in Solanum tuberosum L. plants, and
RT regulation of CDSP 34 expression by ABA and high illumination.";
RL Plant J. 16:257-262(1998).
RN [2]
RP SEQUENCE OF 64-80 AND 274-290, AND SUBCELLULAR LOCATION.
RX STRAIN=cv. Haig;
RX MEDLINE=96352500; PubMed=9717138;
RA Pruvot G., Guine S., Peltier G., Rey P.;
RT "Characterization of a novel drought-induced 34-kDa protein located
RT in the thylakoids of Solanum tuberosum L. plants.";
RL Planta 198:471-479(1996).
CC -!- FUNCTION: Required for normal plant growth. May be both
CC photoprotective and play an ancillary role in photosynthesis. May
CC structurally stabilize thylakoids during osmotic and oxidative
CC stress.
CC -!- SUBUNIT: Associates with the major light-harvesting antenna
CC complex polypeptides of the PSII oxygen-evolving complex (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -!- TISSUE SPECIFICITY: Expressed in leaves.
CC -!- INDUCTION: By high illumination, water stress and abscisic acid.
CC -!- SIMILARITY: BELONGS TO THE LIPC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC EMBL; Y15269; CAA75558.1; -
DR PIR; T07825; T07825.
DR GO; GO:0009535; C:thylakoid membrane (sensu Viridiplantae); IEP.
DR GO; GO:0006950; P:response to stress; IEP.
DR InterPro; IPR006843; PAP-fibrillin.
DR Pfam; PF04755; PAP-fibrillin; 1.
KW Chloroplast; Thylakoid; Membrane; Transit peptide.
FT TRANSIT 1 63 CHLOROPLAST.
FT CHAIN 64 326 LIGHT-INDUCED PROTEIN.
FT CONFLICT 286 286 T -> I (IN REF. 2).
FT SEQUENCE 326 AA; 35635 MW; 1984482198724DEB CRC64;
Query Match 7.4%; Score 78.5; DB 1; Length 326;
Best Local Similarity 25.4%; Pred. No. 33;
Matches 32; Conservative 21; Mismatches 44; Indels 29; Gaps 5;
QY 27 EEINKAIDDAIAAEQSETIDPMKVPDPHADKFERHVGIVDFKGLAMRNIEARGLKOMKR 86
Db EISQITDSFVONS-----VFAGPLATTSISTNAKEVRS 217
QY 87 QGDANVKGEEGIVKAHLIGVHDDIV--SMEYDLYAKGLDHPHTTHVISDIQDFVVALS 143
Db PKRQIKFEEGLIGTQPLT---DSIVLPENVEF-LGQKI-DLSPFKGLITSVQDTASSVA 272
QY 144 LEISDE 149
Db KSISSQ 278
RESULT 80
YE72_HA3IN STANDARD; PRT; 351 AA.
AC P44206;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein H11472 precursor.
GN H11472
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OC NCBI_TaxID=727;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McElroy K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geochagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512 (1995).
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137486; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429 (2000).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

CC or send an email to license@isb-sib.ch.
CC EMBL; U32825; AAC23120.1; -
DR PIR; I64030; I64030.
DR TIGR; H11472; -
DR InterPro; IPR002491; Peripla_BP.
DR Pfam; PF01497; Peripla_BP_2; 1.
KW Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 351 PROTEIN H11472.
FT SEQUENCE 351 AA; 39205 MW; 166842B3F4D2BBED CRC64;
Query Match 7.4%; Score 78.5; DB 1; Length 351;
Best Local Similarity 26.3%; Pred. No. 35;
Matches 44; Conservative 36; Mismatches 54; Indels 33; Gaps 11;
QY 2 MKFLIIA-----AVAFVAVSADPIHYKITBEINKAIDDAI--AAEQSETIDPMKVPDHA 55
Db LKSLIIACLLSLSLFSAL-ADRIITDQDRKV--TIPDHINRAVVLQHOTLN---IAVOL 56
QY 56 DKFERHVGIV-DFKGLAMRNIE-ARGIKOMKRGDANVKGEEGIVKAHLIGVHDDIVS 113
Db DATKQIVGLSNWKKQKGNVRLAPELENMMPGDUN-----SVNIESLLAKPDVVF 110
QY 114 MEYDLAYKGLDHPHTTHV--ISDIQDFVVALSL---EISDEGNITMT 155
Db VT-----NYAPSEMIKQISDVNIPVAISLRTGVEGKGLNPT 149
RESULT 81
LE12_METTH STANDARD; PRT; 505 AA.
AC Q27525;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 2-isopropylmalate synthase 2 (EC 2.3.3.13) (Alpha-isopropylmalate
DE synthase 2) (Alpha-IPM synthetase 2).
GN LEUA2 OR MTH1481.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OC NCBI_TaxID=187420;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keegle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang F., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RT J. Bacteriol. 179:7135-7155 (1997).
RL FUNCTION: Catalyzes the condensation of the acetyl group of
CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC -! CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O =
CC 2-hydroxy-2-isopropylsuccinate + CoA.
CC -! PATHWAY: Leucine biosynthesis; first step.
CC -! SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate
CC synthase family. Leua1 subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

```

-----
CC      EMBL; AB000909; AAB85956.1; -.
DR      PIR; D69064; D69064.
DR      HAMAP; MF_01025; -. 1.
DR      InterPro; IPR002034; AIPM/HcIt synth.
DR      InterPro; IPR000891; HMGL-like.
DR      Pfam; PF00682; HMGL-like; 1.
DR      PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR      PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
KW      Leucine biosynthesis; Transferrase; Complete proteome.
SQ      SEQUENCE 505 AA; 54674 MW; 7DEB80E144C5DA5 CRC64;

Query Match          7.4%; Score 78.5; DB 1; Length 505;
Best Local Similarity 20.6%; Pred. No. 54;
Matches 40; Conservative 36; Mismatches 73; Indels 45; Gaps 8

QY      30  NKAI--DDAIA-----AIQSEITDPM--KVPDHADK--FERHVGIVDFKGLAMR 74
Db      284  NKAIVGENAFAHEAGIHVHGLEAKETVEPTTPMVGHKRRIVLGHKTCANALRSKLQEV 343
QY      75  NTEARG-----LKMKQSGDANV-----KGEIGVKAHLILGVHDDIV 112
Db      344  GLEMKEEQFTYGVKRLGDKGKRITDADRAMAVTILKASREIVKLEGIAVMTGESV 403
QY      113 SMEYDAYLKYGLDHPHTTHV--ISDIQDFWVALSLEISDEGNITMTTFEVRQFANVNHIG 170
Db      404  MPTATVKLRIGDEVXTTSMTGVPVDRAINAIQSLVSETADIELDVNIEAIT-----G 457
QY      171  GLSILDPFGVLSD 184
Db      458  GTNALAEVFWMSD 471

RESULT 82
YAC8 YEAST
ID      YAC8 YEAST STANDARD; PRT; 528 AA.
AC      P39734; P39733;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Hypothetical 58.8 kDa protein in MYO4-DS2 intergenic region.
GN      YAL028W
OR      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / AB972;
RX      MEDLINE=95549563; PubMed=7731988;
RA      Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA      Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA      Storms R.K.;
RA      "The nucleotide sequence of chromosome I from Saccharomyces
RT      cerevisiae."
RL      Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813 (1995).
RN      [2]
RP      REVISIONS.
RA      Vo D.T.;
RC      Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC      or send an email to license@isb-sib.ch)
CC
CC      EMBL; U12980; AAC05004.1; -.
DR      PIR; S70295; S70295.
DR      GERMOnline; 138370; -.
DR      SGD; S0000026; YAL028W.
-----

```

```

Matches 44; Conservative 43; Mismatches 91; Indels 49; Gaps 7;
QY 13 VAVSADPIHYDKITEEINKAIDDAIAEQSETIDPMKVPDPHADKFERHVGIVDFKGLA 72
Db 107 VAAGMNPX---DLKRGIDKAVDSAV-----EKLRAQAQCSKESITQVGSISANSRA 157
QY 73 MRNTEARGLKQMKRQGDANVKGEGIVKAHLIG-----VHDDIVSMEYDL 118
Db 158 IGDIIAEMEKVGRNGVITVEEGQGLSNELSVVEGQFDRGYLSFYFINN 217
QY 119 AY-----KLGLDHPHTHVSIDQFVVALSLEISDEGNITMTSFEVQRQANVVNHI 169
Db 218 PYILLVDKVSIRELLPVLESVAKSRSLIIAEDIDGALATL-----VNNL 267
QY 170 GGL-----SILDPIFGVLSVLTAFQDVTVRKMTKVLAPAFKRELEK 212
Db 268 RGIVRAAAVAPGPG---DNKAMLEDAVLTAGTVISEBIGLELEK 311

RESULT 84
CH60_XANCH STANDARD; PRT; 546 AA.
ID CH60_XANCH Q8RIT7;
AC Q8RIT7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL.
OS Xanthomonas campestris (pv. phaseoli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=29445;
RN [1]
RP SEQUENCE FROM N.A.
RA Hsu C.-C., Yu Y.-J., Yang M.-T.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF426387; AAL74150.1; -
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaperin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC SEQUENCE 546 AA; 57225 MW; 9DDAB51DED42BA5C CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 546;
Best Local Similarity 20.4%; Pred. No. 59;
Matches 49; Conservative 41; Mismatches 79; Indels 71; Gaps 11;
QY 13 VAVSADPIHYDKITEEINKAIDDAIAEQSETIDPMKVPDPHADKFERHVGIVDFKGL 71
Db 107 VAAGMNPX-----DLKRGIDQAVKAAVVELKNISK---PTTDDKAIAQVGTISANSDE 156
QY 72 AMRNIEARGLKQMKRQGDANVKGEGIVKAHLIGVH-----DDIVSMEYDLAY 120

```

```

Db 157 SIGNIIAEMQKV-----GKEGVITVEGSGLENELDVWEGMDFRGYLSFYFINN 207
QY 121 -----KLGLDHPHTH--VISIDQFVVAL-----SLEISDEGNITMTSFEV 159
Db 208 QOSQASDLDPPFILLHDKYKISNVRDLIPVLEGVAKGAKGPLLIVAEVEGALATL----- 262
QY 160 RQFANVVNHIGL-----SILDPFEG-----VLSVLTAFQDVTVRKMTKVLAPAFKREL 210
Db 263 -----VNTIRGIVKVAAPGFGDRRKAMLEDMAVLTGTVISERVGLAEKATIKDL 317

RESULT 85
CH60_COLMA STANDARD; PRT; 549 AA.
ID CH60_COLMA Q93GT8;
AC Q93GT8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein).
GN GROEL OR GROEL.
OS Colwellia maris (Vibrio sp. (strain ABE-1)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Colwellia.
OX NCBI_TaxID=77524;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22883852; PubMed=12898133;
RA Yamauchi S., Okuyama H., Morita E.H., Hayashi H.;
RT "Gene structure and transcriptional regulation specific to the groESL
RT operon from the psychrophilic bacterium Colwellia maris.";
RL Arch. Microbiol. 180:272-278(2003).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB073221; BAB70476.2; -
CC HAMAP; MF_00600; -; 1.
CC InterPro; IPR001844; Chaperin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding.
CC SEQUENCE 549 AA; 57690 MW; 3A194880966062A2 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 549;
Best Local Similarity 22.1%; Pred. No. 60;
Matches 52; Conservative 36; Mismatches 82; Indels 65; Gaps 11;
QY 13 VAVSADPIHYDKITEEINKAIDDAIAEQSETIDPMKVPDPHADKFERHVGIVDFKGLA 72
Db 107 IAAGMNPX---DLKRGIDKAVIAAVALKESST-----FVTGNKAEQVGTISANSDET 157
QY 73 MRNIEARGLKQMKRQGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKLGDLHP----- 127
Db 158 VGKIIATAMEKVGEGVITV---EAG-----QALTDELVDVWEGMDFRGY-----LSFYFINN 207
QY 128 -----TTHVISIDQFVVALS-----LEISD---EGNITMTSFEVQR 161

```

```
Db 208 QENGTVLENPPFILLVKKISNIRELLTLEGVAKAGKPELLIIAEDVEGEALATL-----262
Qy 162 FANVNHIGGL-----SILDPFGVLSVDLTATFQDTRKEMTKVLAPAFKRELEX 212
Db 263 ---VNNMRGIVKVAAPGFG---DRRKAMLDQVATLTAGTVISEIGMELEK 311

RESULT 86
LON CHLMU STANDARD; PRT; 819 AA.
AC QPKS6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR TC0623
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: Degrades short-lived regulatory and abnormal proteins in
CC presence of ATP. Hydrolyzes two ATPs for each peptide bond cleaved
CC in the protein substrate (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002331; AAF39454.1; -.
CC PIR; B81681; B81681.
CC MEROPS; S16.002; -.
CC TIGR; TC0623; -.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR001270; Chaprinin_c1pA/B.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_Lon.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008268; Peptid_S16_AS.
CC InterPro; IPR001984; Peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02130; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00300; CLPPROTEASEA.
CC PRINTS; PR00830; ENDOLAPTAASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC TIGRFAMs; TIGR00763; Lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
CC Hydrolase; Serine protease; ATP-binding; Complete proteome.
KW NP_BIND 392 399
FT ACT_SITE 724 724
FT BY SIMILARITY.
```

```
SQ SEQUENCE 819 AA; 91843 MW; DD87D5D701E1B063 CRC64;
Query Match 7.4%; Score 78.5; DB 1; Length 819;
Best Local Similarity 21.8%; Pred. No. 95;
Matches 51; Conservative 38; Mismatches 78; Indels 67; Gaps 11;

Qy 16 SADPIHYDKITEINKAIDDAIAAIBOSEITDP-----MKVP-----52
Db 6 NTDSONLDPNASEVEKLLDSEAEAEKTDHTPPSELFLPNKRPPFGMAAPLLIEAG 65
Qy 53 DH-----ADKPERHYGIVDFKGLAMRNIEARGLKQMKRQGDANVKYEGEIVKAHLII 105
Db 66 PHYEVLTLLAKSSQKHGLVLTKKEDA--NTLKIGFNLHRVG-----VSARIL--112
Qy 106 GVHDDIVSMYEDLAYKLGDLHPHTTHVISDIDQFVVALSLEISDEGNITMTSFVRQPA-N 164
Db 113 ---RIMPIEGGSAQVLLSIEDIRIVKVPQDKYLKAKVAYHKE-NKELTE-ELKAYSIS 166
Qy 165 VVNHIGLSILDPF-----GVLSVDLTATFQDTRKEMTKVL 202
Db 167 IVSIIKDLLKLNLPLFKEELQIFLGHSDFTPEPKLADPSVALTTAT-REELQEVLL 219

RESULT 87
SYA METJA STANDARD; PRT; 892 AA.
AC Q57984;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
GN ALAS OR MJ0564.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=6688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RT Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67505; AAB98556.1; -.
CC PIR; D64370; D64370.
CC TIGR; MT0564; -.
CC HAMAP; MF_00036; -.
CC InterPro; IPR003156; DEHAL.
CC InterPro; IPR002318; tRNA-synt_2c.
CC InterPro; IPR006193; tRNA_synt_Ala.
CC Pfam; PF02272; DEHAL; 1.
CC Pfam; PF01411; tRNA-synt_2c; 1.
```



```

RC RC STRAIN=cv. Columbia;
RR Despresz T., Anselme J., Chiapello H., Rouze P., Caboche M., Hofte H.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
[3]
RRP SEQUENCE OF 108-191 FROM N.A.
RC STRAIN=cv. Columbia;
RR Berthomieu P., Guerrier D., Giraudat J.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
[4]
RRP SEQUENCE OF 197-407 FROM N.A.
RC STRAIN=cv. Columbia;
RR Raynal M., Grillet F., Laudie M., Meyer Y., Cooke R., Delsey M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: Belongs to the L4E family of ribosomal proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AL162973; CAB86041.1; -.
DR EMBL; Z18118; CAA79104.1; -.
DR EMBL; Z18460; CAA79191.1; -.
DR EMBL; Z17589; CAA79004.1; -.
DR PIR; T48308; T48308.
DR InterPro; IPR002136; Ribosomal L4/L1E.
DR Pfam; PF00573; Ribosomal L4; 1
DR PROSITE; PS00939; RIBOSOMAL_L1E; 1.
DR Ribosomal protein.
KW CONFLICT 181 G -> V (IN REF. 3).
FT CONFLICT 362 E -> D (IN REF. 4).
SQ SEQUENCE 407 AA; 44722 MW; 0266FC2852B7A2038 CRC64;
-----
Query Match 7.3%; Score 78; DB 1; Length 407;
Best Local Similarity 26.4%; Pred. No. 46;
Matches 29; Conservative 24; Mismatches 41; Indels 16; Gaps 4;

QY 1 MMKFLLLIAAVFAVSGADPT---HYDKITEEINKAIDDAIAAEQS----ETIDPMKVPD 53
Db 125 MKRHAIVSAIAATAVPALVMARGHKTIENVPENPLVVSDESAEAKTSAAIKVLKIGAYD 184
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 54 HADYFERHVGIVDPKGLAMNTEARGLQKMGKQGDVANKGEEG--IVKA 101
Db 185 DAERKAKNSIGIRPKGKGV-----RNRRYIRKKGELVVFGTEGAKIVKA 227
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 91
CH60_R1CTS
ID CH60_R1CTS STANDARD; PRT; 555 AA.
AC F16625;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Major antigen 58)
DE (58 kDa antigen).
DE GROEL OR GROEL OR MODA OR STA58.
OS Rickettsia tsutsugamushi.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Orientia.
OX NCBI_TaxID=784;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=90216005; PubMed=2108930;
RR Stover C.K., Marana D.P., Dasch G.A., Oaks E.V.;
RT "Molecular cloning and sequence analysis of the Sta58 major antigen
RT gene of Rickettsia tsutsugamushi: sequence homology and antigenic
RT comparison of Sta58 to the 60-kilodalton family of stress proteins.";
RL Infect. Immun. 59:1360-1368(1990).
CC -! FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress

```

CC conditions (By similarity).
 CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
 CC 7 subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M31887; AAA26393.1; -.
 CC PIR; B41492; B41492.
 CC HSSP; P06139; 1GRL.
 CC HAMAP; MF_00600; -; 1.
 CC InterPro; IPR001844; Chaperonin_Cpn60.
 CC InterPro; IPR002423; Cpn60/TCP-1.
 CC InterPro; IPR008950; GroEL-ATPase.
 CC Pfam; PF00118; cpn60_TCP1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PROSITE; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 CC Chaperone; ATP-binding; Antigen.
 CC SEQUENCE 555 AA; 59729 MW; 16400249D3FC03B4 CRC64;

Query Match 7.3%; Score 78; DB 1; Length 555;
 Best Local Similarity 25.3%; Pred. No. 66;
 Matches 40; Conservative 19; Mismatches 77; Indels 22; Gaps 4;
 QY 18 DPHYDKITEINKAIDDATAAIEQSETIDPMKVPDHDKFERHVGIVDFKGLAMRNIE 77
 DB 401 DALRATRAVDEEIVPGGGVAFVSRVLSLKP-----DNEDORVGI-----NII 446
 QY 78 ARGILKMKRQGDANVKEEGIVKAHLIGVHD-----DIVSMYDLYAKLGLDHPHTTHVIS 133
 DB 447 KKVLEAPVRQIVKQAGKEDVNVNELSKSTDKNRGPDAWQVDMIKAGIVPTKVVRT 506
 QY 134 DIQDFVVALSLEISDEGNITWTFSEVQFANVNVHIG 171
 DB 507 ALQDAFVSFVLIATSAMIT-----DHEDNNTGNRSGG 540

RESULT 92
 Y4FB RHISN
 ID Y4FB RHISN STANDARD; PRT; 664 AA.
 AC P55440;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 73.7 kDa protein Y4FB.
 GN Y4FB.
 OS Rhizobium sp. (strain NGR234).
 OC Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 OX NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed=9163424;
 RA Freiberg C.A., Fellay R., Baroch A., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL Nature 387:394-401(1997).
 CC -!- SIMILARITY: None obvious.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000072; AAB91659.1; -.
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001604; Endonuclease.
 CC Pfam; PF01223; Endonuclease; 1.
 CC SMART; SM00477; NUC; 1.
 CC KW Hypothetical protein; Plasmid.
 CC SEQUENCE 664 AA; 73731 MW; CFCC041FB73C064F CRC64;
 QY
 Query Match 7.3%; Score 78; DB 1; Length 664;
 Best Local Similarity 22.6%; Pred. No. 82;
 Matches 45; Conservative 27; Mismatches 65; Indels 62; Gaps 9;
 QY 28 EINKAIDDAI-----AAIEQSETIDPMKVPDHA-----DKFERHVGIVDFKGE 70
 DB 2 EISKFADEQLADQIAAGREAVPEVDPIRLIQVVRARAGAGVDFVFESLIGED---- 57
 QY 71 LAMRNIEARGLKQMKR-----QGDANVKEEGIVKAHL-----IGVHDDIVSME 115
 DB 58 LVEINYLERGLMASRAVCINVPAPVGSSDWGTGFLIGPRLLLNNHVIDSAEDALKAT 117
 QY 116 YDLAYKL---GDLHPTTH-----VIS--DIQDFVVALSLEISDEGNITWTF---- 157
 DB 118 VEFVGYELDAEGELKKTTRFLRSLPQDGFVTSRDLADYTVVAIENSEDGTATPISDFGLR 177
 QY 158 -----EVQFANVNVH 168
 DB 178 LDARTGKTEVGQYATIIGH 196
 RESULT 93
 POLS_IBDV5
 ID POLS_IBDV5 STANDARD; PRT; 1012 AA.
 AC P25219;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Structural polypeptide [Contains: Major structural protein VP2;
 DE Nonstructural protein VP4; Minor structural protein VP3].
 OS Avian infectious bursal disease virus (strain 52/70) (IBDV).
 OC Viruses; dsRNA viruses; Birnaviridae; Avibirnavirus.
 OX NCBI_TaxID=10996;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90278420; PubMed=2161902;
 RA Bayliss C.D., Spies U., Shaw K., Peters R.W., Papageorgiou A.,
 RA Mueller H., Boursnell M.E.G.;
 RT "A comparison of the sequences of segment A of four infectious bursal
 RT disease virus strains and identification of a variable region in
 RT VP2.";
 RL J. Gen. Virol. 71:1303-1312(1990).
 CC -!- FUNCTION: Segment A encodes a polypeptide that is processed into
 CC the major structural proteins of the virion VP2 and VP3, and into
 CC the putative protease VP4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00869; BAA00745.1; -.
 CC PIR; JQ0941; GNXS52.
 CC MEROPS; S50.002; -.
 CC InterPro; IPR002662; Birna_VP2.
 CC InterPro; IPR002663; Birna_VP3.
 CC InterPro; IPR002664; Peptidase_S50.
 CC Pfam; PF01766; Birna_VP2; 1.
 CC Pfam; PF01767; Birna_VP3; 1.


```

CC  -!- FUNCTION: May play a role in the repair of endogenous alkylation
CC  damage (By similarity).
CC  -!- SIMILARITY: Belongs to the recA family. RadA subfamily.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL curation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D26185; BAA05321.1; -;
DR  EMBL; Z99104; CAB11863.1; -;
DR  EMBL; U02604; AAA19234.1; -;
DR  PIR; S66116; S66116.
DR  Subtilist; BG10149; radA.
DR  InterPro; IPR0031593; AAA_ATPase.
DR  InterPro; IPR004504; PepC_S16_RadA.
DR  SMART; SM00382; AAA; 1.
DR  TIGRfams; TIGR00416; sms; 1.
KW  DNA repair; ATP-binding; Zinc-finger; DNA-binding; Complete proteome.
FT  ZN FING 10 27 C4-TYPE (POTENTIAL).
FT  NP BIND 98 105 ATP (POTENTIAL).
SQ  SEQUENCE 458 AA; 49483 MW; 7397601CF42C801B CRC64;
-----
Query Match 7.3%; Score 77.5; DB 1; Length 458;
Best Local Similarity 22.5%; Pred. No. 58;
Matches 40; Conservative 23; Mismatches 56; Indels 59; Gaps 5;
-----
QY 25 ITBEINKAIDDAIAAEQSS-----ETIDPMKVPDHADKFERHVGIVDFKGSL 71
Db 34 VEEMIKKAPNRRAPSHSVQIVKQPSPTSIETSEPRVKTLGEPNVLGGVVKGSL 93
QY 72 AMRNIETARGLKQMKRGDANVKGEGIVKAHLIIGVHDDIVSMEDVLA- 120
Db 94 VL-----IGGDPGKSTLLQVSAQLSGSSNSVLVYISGEESVKQTK 135
QY 121 ----KLGDLPHTTHVSD-----IQD----FVVALSLRISDEGNITMTSFVRQ 161
Db 136 LRADRLGNNPNSHLVLSSETDMYVISAQEMNPFVVDVSIQTVIQSDITSAPGSVQ 193
-----
RESULT 97
CH60_ENTIT STANDARD; PRT; 539 AA.
AC O66132;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaparonin (Protein Cpn60) (groEL protein) (Fragment).
DE GROEL OR GROEL OR MOPA.
OS Enterobacter intermedius.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_Taxid=61648;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=JCM 1238;
RA Harada H., Ishikawa H.;
RT "Phylogenetical relationship based on groE genes among phenotypically
RT related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT species."
RL J. Gen. Appl. Microbiol. 43:355-361(1997).
CC  -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC  proper assembly of unfolded polypeptides generated under stress
CC  conditions (By similarity).
CC  -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC  7 subunits (By similarity).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC  -!- SIMILARITY: Belongs to the chaparonin (HSP60) family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration

```


Search completed: August 6, 2004, 16:00:53
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 15:57:51 ; Search time 38 Seconds
(without alignments)
1768.563 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 1068

Sequence: 1 MMKFLIIAAVAFVAVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	109	10.2	568	10 Q9LXF5	Q9LXF5 arabidopsis
2	103	9.6	436	2 Q9RNV5	Q9RNV5 streptococc
3	103	9.6	436	16 Q8DS90	Q8DS90 streptococc
4	100.5	9.4	531	5 Q8I2H1	Q8I2H1 plasmodium
5	99	9.3	260	5 Q9V3A0	Q9V3A0 drosophila
6	97.5	9.1	776	16 Q98PT2	Q98PT2 mycoplasma
7	96.5	9.0	544	16 Q8FNV8	Q8FNV8 corynebacte
8	96	9.0	234	4 Q8N4C7	Q8N4C7 homo sapien
9	96	9.0	959	16 Q8LWA0	Q8LWA0 bacillus an
10	95.5	8.9	186	2 Q8KVC7	Q8KVC7 uncultured
11	95	8.9	436	16 Q8E3T9	Q8E3T9 streptococc
12	93.5	8.8	449	17 Q9HR60	Q9HR60 halobacteri
13	93.5	8.8	628	17 Q9V1N1	Q9V1N1 pyrococcus
14	93	8.7	436	16 Q8DY73	Q8DY73 streptococc
15	92.5	8.7	637	10 Q8XIJ5	Q8XIJ5 arabidopsis
16	92.5	8.7	859	16 Q7VJY3	Q7VJY3 helicobacte

17	92	8.6	186	2 Q8KVM3	Q8KVM3 uncultured
18	92	8.6	186	2 Q8KV02	Q8KV02 uncultured
19	92	8.6	186	2 Q8KHA5	Q8KHA5 uncultured
20	91.5	8.6	979	16 Q98PK8	Q98PK8 mycoplasma
21	91	8.5	186	2 Q8KV56	Q8KV56 uncultured
22	91	8.5	186	2 Q8KVJ7	Q8KVJ7 uncultured
23	91	8.5	444	2 Q84H55	Q84H55 anaplasma m
24	90.5	8.5	186	2 Q8KV52	Q8KV52 uncultured
25	90.5	8.5	681	5 Q8I126	Q8I126 caenorhabdi
26	90	8.4	186	2 Q8KVE0	Q8KVE0 uncultured
27	90	8.4	186	2 Q8KVE1	Q8KVE1 uncultured
28	90	8.4	333	16 Q8DD69	Q8DD69 vibrio vuln
29	90	8.4	1558	5 Q96275	Q96275 plasmodium
30	90	8.4	1786	5 Q9U0P0	Q9U0P0 plasmodium
31	89.5	8.4	186	2 Q8KVL4	Q8KVL4 uncultured
32	89.5	8.4	939	2 Q8KWS5	Q8KWS5 marionomas
33	88.5	8.3	279	16 Q7MNV1	Q7MNV1 bordetella
34	88.5	8.3	279	16 Q7W160	Q7W160 bordetella
35	88.5	8.3	1393	16 Q822J2	Q822J2 chlamydomophi
36	88.5	8.3	2137	10 Q9SVN7	Q9SVN7 arabidopsis
37	88	8.2	609	16 Q8PGF6	Q8PGF6 escherichia
38	87.5	8.2	485	16 Q897K6	Q897K6 clostridium
39	87.5	8.2	1620	5 Q8IK83	Q8IK83 plasmodium
40	87	8.1	186	2 Q8KV75	Q8KV75 uncultured
41	87	8.1	372	10 Q9ZVB0	Q9ZVB0 arabidopsis
42	87	8.1	385	10 Q8GWW1	Q8GWW1 arabidopsis
43	87	8.1	399	5 Q4J999	Q4J999 caenorhabdi
44	87	8.1	562	17 Q8U2M6	Q8U2M6 pyrococcus
45	87	8.1	1640	9 Q9AZS0	Q9AZS0 bacterioph
46	87	8.1	1640	16 Q9CF87	Q9CF87 lactococcus
47	86	8.1	280	16 Q7TZ25	Q7TZ25 mycobacteri
48	86	8.1	303	16 Q9I432	Q9I432 pseudomonas
49	86	8.1	339	16 P71763	P71763 mycobacteri
50	86	8.1	1661	5 Q06166	Q06166 plasmodium
51	85.5	8.0	186	2 Q8KV93	Q8KV93 uncultured
52	85.5	8.0	186	2 Q8KVA8	Q8KVA8 uncultured
53	85.5	8.0	263	4 Q86YC9	Q86YC9 homo sapien
54	85.5	8.0	264	9 Q7Y444	Q7Y444 enterobacte
55	85.5	8.0	461	16 Q7VFQ7	Q7VFQ7 helicobacte
56	85.5	8.0	745	4 Q8N3E4	Q8N3E4 homo sapien
57	85.5	8.0	807	16 Q7UYX3	Q7UYX3 rhodopirell
58	85.5	8.0	990	16 Q8EAY1	Q8EAY1 shewanella
59	85.5	8.0	1062	4 Q7Z619	Q7Z619 homo sapien
60	85.5	8.0	1154	11 Q91ZS0	Q91ZS0 mus musculu
61	85.5	8.0	1441	4 Q60310	Q60310 homo sapien
62	85.5	8.0	2238	11 Q91ZS1	Q91ZS1 mus musculu
63	85.5	8.0	2780	11 Q91ZS2	Q91ZS2 mus musculu
64	85.5	8.0	6298	11 Q8VHN7	Q8VHN7 mus musculu
65	85	8.0	185	-2 Q8KVK6	Q8KVK6 uncultured
66	85	8.0	186	2 Q8KVC0	Q8KVC0 uncultured
67	85	8.0	186	2 Q8KV82	Q8KV82 uncultured
68	85	8.0	186	2 Q8KHQ8	Q8KHQ8 uncultured
69	85	8.0	292	17 Q9UZS6	Q9UZS6 pyrococcus
70	85	8.0	401	16 Q8RDZ2	Q8RDZ2 fusobacteri
71	85	8.0	537	16 Q8DBL1	Q8DBL1 vibrio vuln
72	85	8.0	570	16 Q8FO17	Q8FO17 leptospira
73	85	8.0	878	16 Q9CFU9	Q9CFU9 lactococcus
74	84.5	7.9	334	2 Q8RQU4	Q8RQU4 thermus the
75	84.5	7.9	426	16 Q8AOK9	Q8AOK9 bacteroides
76	84.5	7.9	710	16 Q8PG51	Q8PG51 xanthomonas
77	84.5	7.9	817	16 Q8CQ88	Q8CQ88 staphylococ
78	84.5	7.9	841	16 Q898W1	Q898W1 clostridium
79	84.5	7.9	1429	5 Q8SXM3	Q8SXM3 drosophila
80	84	7.9	186	2 Q8KVI9	Q8KVI9 uncultured
81	84	7.9	186	2 Q8KVI5	Q8KVI5 uncultured
82	84	7.9	306	16 Q8DWU5	Q8DWU5 streptococc
83	84	7.9	314	16 Q9WZQ7	Q9WZQ7 thermotoga
84	84	7.9	392	16 Q99X76	Q99X76 staphylococ
85	84	7.9	600	2 Q9X916	Q9X916 versinia en
86	84	7.9	600	16 Q92375	Q92375 versinia pe
87	84	7.9	775	16 Q9CHZ3	Q9CHZ3 lactococcus
88	84	7.9	9510	2 Q93NX9	Q93NX9 streptomyce
89	83.5	7.8	186	2 Q8KVH7	Q8KVH7 uncultured

90 83.5 7.8 266 16 Q7VTQ1
91 83.5 7.8 311 2 Q93R43
92 83.5 7.8 311 2 Q93R44
93 83.5 7.8 311 2 Q93R44
94 83.5 7.8 329 2 Q9EXN7
95 83.5 7.8 402 11 Q8K3U8
96 83.5 7.8 412 16 Q8NUB6
97 83.5 7.8 474 16 Q97K37
98 83.5 7.8 1047 4 Q9C0B4
99 83.5 7.8 1302 4 Q7Z2L3
100 83 7.8 186 2 Q8KVP2

ALIGNMENTS

RESULT 1
Q9LXFS PRELIMINARY; PRT; 568 AA.
AC Q9LXFS
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN F8W21_160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 4 KH DOMAINS.
DR EMBL; AL353993; CAB89337.1; -.
DR PIR; T49962; T49962.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH; 4.
DR SMART; SMO0322; KH; 4.
DR PROSITE; PS00084; KH_TYPE_1; 4.
KW Hypothetical protein.
SQ SEQUENCE 568 AA; 62034 MW; 3431FB92B4E37205 CRC64;

Query Match 10.2%; Score 109; DB 10; Length 568;
Best Local Similarity 26.7%; Pred. No. 1.5;
Matches 46; Conservative 36; Mismatches 64; Indels 26; Gaps 8;
QY 25 ITEBKAIIDAAIAEIQSEITDPMK--VPDHADKFERHVGIVDFKGLAMRN-----IE 77
DB 331 ITISAREVFDA----YSPTEIAVMRLQPKCDKVERDGLSVFTTLLVPPSSRGICL 385
QY 78 ARG---LKQMKRGQDANVK--GEGIVKA-----HLLIGVHDDIVSMYDYLAYKGLDHP 128
DB 386 GKGGAIITETRMKATIRILGKENVLPKVASDDDDMMQVNFVWMLLLKFSQFLLSRLRYF 445
QY 129 THVTSIDIDQFVVALS--LEISDEGNITWTSPEVRFQFANVNHIGGLSILDP 178
DB 446 TWSVNSNSNFVQVLSGELDVAKELIQITS---RLRANVPDEGAVSALMEV 494

RESULT 2
Q9RHVS PRELIMINARY; PRT; 436 AA.
AC Q9RHVS
DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Phosphoglycerate dehydrogenase.
GN PGDA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT8149.
RA Kawabata S., Teraso Y., Hamada S.;
RT "Molecular cloning, sequence and characterization of a novel
RT streptococcal phosphoglycerate dehydrogenase gene.";
RL Oral Microbiol. Immunol. 15:58-62 (2001).
DR EMBL; AB016077; BAA88823.1; -.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR006073; GTP_OBG.
DR InterPro; IPR002917; MMR_HSR1.
DR Pfam; PF01926; MMR_HSR1; 1.
DR PRINTS; PR00326; GTP_OBG.
DR TIGRFAMs; TIGR00650; MG442; 2.
DR TIGRFAMs; TIGR00231; small_GTP; 2.
SQ SEQUENCE 436 AA; 48601 MW; EC52CC8650DED090 CRC64;
Query Match 9.6%; Score 103; DB 2; Length 436;
Best Local Similarity 23.0%; Pred. No. 3.2;
Matches 50; Conservative 36; Mismatches 77; Indels 54; Gaps 11;
QY 30 NKAIDDAIAAEIQSEITDPMKVPDHADKFERHVGIVDFKGLAMRNIEARGLKQMKRQGD 89
DB 21 NRIAGERISIVEDVEGVTRDRIYTKAEWLNRFSLIDTGG---IDVDAPFMEQIKHQAD 77
QY 90 AN-----VRGEGIVK-----AHLIGVHDDIV-----SMEYDL-AYK 121
DB 78 IAMTEADVIVFVSAKEGITDADEVAKILYRTHKPVILAVNKVDNPEMSAIYDFVALG 137
QY 122 LGDLHP--TTHVI--SDIQDFVW-ALSLIEISDEGNITWTSPEVRF-----ANVV 166
DB 138 LGDPYFVSSAHGIGTGVDLDAIVNLPAAQEE-----SSDIKFLSLIGPNVYKSSLI 191
QY 167 NHIGLS--ILDPIFGLSDVLTATFQDTVKEMTKV 201
DB 192 NAILGEDRVIASPVAGTTRDAITDTFTDEGEQETMI 228
RESULT 3
Q8DS90 PRELIMINARY; PRT; 436 AA.
AC Q8DS90
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Phosphoglycerate dehydrogenase.
GN PGDA OR SMU.1920.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE015016; AAN59531.1; -.
DR InterPro; IPR005289; GTP-binding_dom.
DR InterPro; IPR006073; GTP_OBG.

DR InterPro; IPR002917; MMR_HSR1.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01926; MMR_HSR1; 1.
DR PRINTS; PR00326; GTP1OBG.
DR TIGRFAMs; TIGR00650; MG442; 2.
DR TIGRFAMs; TIGR00650; MG442; 2.
KW Complete proteome.
SQ SEQUENCE 436 AA; 48585 MW; 24D842BA91C2A097 CRC64;

Query Match
Best Local Similarity 9.6%; Score 103; DB 16; Length 436;
Matches 50; Conservative 36; Mismatches 77; Indels 54; Gaps 11;

QY 30 NKAIDDAIALEQSETIPMKVDPHADKFERHVGIVDFKGELARNTEARGLKOMKEQG 89
DB :
21 NRAGERISIVEDVEGVTRORIYTKAEWLNRQFSIIDTGG---IDDVDAPPMEIQKHQA 77
DY :
QY 90 AN-----VKGEEGIVK-----AHLLTGIVDDIV-----SMEDYL-AYK 121
DB :
78 IAMTEADVIWVVSAKEGITDADEVAKILRYTHKPVLAVKNKYDNPEMSAIYDFVALG 137
DY :
QY 122 LGDLHP--THVI--SDIQPFW-ALSLEISDEGNITWTSEVRQF-----ANNV 166
DB :
138 LGGDPYPVSSAHGIGTGVDLDAIVNLPTEAQEE-----SSDIKKFSLGRPNVNGKSII 191
DY :
QY 167 NHIGGLS-IILDPIFGVLSLVLTAFODTVRKEMTKV 201
DB :
192 NAILLGEDRVIASPVAGITRDATDTTFDEEQEFTMI 228
DY :

RESULT 4

Q8I2H1 PRELIMINARY; PRT; 531 AA.

AC Q8I2H1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF11680W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
NCBI_TaxId=36329;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Gobie A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitzsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR ENBL; AL929359; CAD52022.1; -
DR InterPro; IPR006577; UNS.
DR InterPro; IPR001012; UBX.
DR Pfam; PF00789; UBX; 1.
DR SMART; SMOO594; UAS; 1.
DR PROSITE; PS50033; UBX; 1.
KW Hypothetical protein.
SQ SEQUENCE 531 AA; 62569 MW; 7D1867D7919C7A11 CRC64;

Query Match
Best Local Similarity 9.4%; Score 100.5; DB 5; Length 531;
Pred. No. 6.8;

RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RA MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazer R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Teang G., Wan K., Whitelaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazer R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.B., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lonoten M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.N., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AE003641; AAF53345.1; -;
DR EMBL; AE003408; AAF44828.1; -;
DR Flybase; FBGN0028920; BG:DS00941.12.
DR InterPro; IPR004272; Odorant_binding.
DR Pfam; PF03027; DUF233; 1.
DR SMART; SM00700; JHBP; 1.
KW Hypothetical protein.
SQ SEQUENCE 260 AA; 28095 MW; FC3486D66A29A6AB CRC64;

Query Match 9.3%; Score 99; DB 5; Length 260;
Best Local Similarity 21.9%; Pred. No. 3.5;
Matches 56; Conservative 41; Mismatches 63; Indels 96; Gaps 13;

QY 5 LLIAAVFVAVSA-----DPHYDKITEINKAIDDAIAEQS-----ETIDPM 49
DB 3 IFVAILAFVAVSAASNGQPIET---TQSSITVDVIEGIEKQPCGFTSVGLPLAPL 58
QY 50 KVPDHAKFERHVGIVDFKGLAENVNTEARGLKQKQGDANVKGEGIVAKHLLIGVHD 109
DB 59 RI-DHQD-----INI-----DSSVLKQAQGTIDHFLNGLND 88
QY 110 -DI-----VSMVEYDLAYKGLDLPHTTHVIS-----DIQDFV 140
DB 89 FDIDEMKNATSKVTKYKFTFRDNNVDQDLSVLLKKYGTINLIGAGHAKFAIKDWI 148
QY 141 ALSLEIS---DEGNTMTSFEVRFQFANVNVHIGLSILDPIFGVLSVLFAIQDVTYKE 197
DB 149 WGTWKYSGLVSGNKLKSLVR-----TLHG--EVDSEIEGILGD-----GSINEK 193
QY 198 MTKVLAPAFKRELEKN 213
DB 194 MNEYLAEAVALAINEN 209

RESULT 6
Q98PT2 PRELIMINARY; PRT; 776 AA.
AC Q98PT2
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE LIPOPROTEIN.
GN MYPU 6370.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445565; CAC13810.1; -;
DR PIR; E90591; E90591.
DR Mypulist; MYPU 6370; -;
DR InterPro; IPR00437; Prok_lipoprot_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 776 AA; 88710 MW; 94F1B4A485FA4DB4 CRC64;

Query Match 9.1%; Score 97.5; DB 16; Length 776;
Best Local Similarity 20.1%; Pred. No. 20;
Matches 52; Conservative 37; Mismatches 73; Indels 97; Gaps 10;

QY 22 YDKITEINKAIDDAIA---AIEQSETIDPMKVPD-HADKFE-----RHVGIVD--- 66
DB 437 YNQRFDQVNLKLEQAFPKNSWEQTNPSPDESNTYQDKLDNIGDEDEKAKFIGITDLP 496
QY 67 -----FKGELAMRN-----IEARGLKQKQGDANVKGEGIVK-----A 101
DB 497 NELKSLEFKKAIALRNAIAAIIGLNTLAELKKGKGSQWASRDILVKADFEKEMQHDNDNA 556
QY 102 HL-----LIGVHDDIVSMVEYDLAYKGLDLPHTTHVISDIQDFVVALSLEIS----- 147
DB 557 HLRKTEKYIFGVNGDEQSDKDLVKQLSKQSFISAVESIH---AKALESRKYAKNV 612
QY 148 DEGNI-----TMTSFEVRFQFANVNVH 168
DB 613 DENNIKAVELYLANYNQAIWNKFKNLYEIKWTKSKFKKDGSGSSMTIEQVINEYVK 672
QY 169 IGLSILDPFGVLSVL 187
DB 673 YRSSNVADKLIISVITSTNT 691

RESULT 7
Q98NV8 PRELIMINARY; PRT; 544 AA.
AC Q98NV8
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter ATP-binding protein.
GN CE2035.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kwarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF005221; BAC18845.1; -;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.

```
DR PROSITE; PS00211; ABC TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Complete; Proteome
SQ SEQUENCE 544 AA; 59593 MW; 3BE49CA60450A53E CRC64;

Query Match
Best Local Similarity 24.5%; Score 96.5; DB 16; Length 544;
Matches 47; Conservative 30; Mismatches 64; Indels 51; Gaps 10;

QY 19 PIHYDKITEE-----INKAIDDAIAAEQSETI-----DPMKVPDHAKFERH 61
Db 67 PETHGAGTDDPFDLPVSLIDEAVPRAVETIGELTVMGDDNPPPIAERFDBA 126
QY 62 VGIYDFKGLAMENIAR-----GLKQKQGDANYKGBEG-----IVKAHLILGVH 108
Db 127 LSRAE---ELGVNLEARIIEIIVAGLGLADVDR--SVPIRALSGGQRRRFALAVILLEPH 181
QY 109 DDIVSMEYDLAYKGLDHPHTHVISDIQDFVALSLEISDEGNITMTSEVRFQFANVNH 168
Db 182 DAILDE-----PTNHLDDNAVDFLI--NELTSFKGPVLTAHSD--RFFLDVAV-- 225
QY 169 IGGSLIDPIFG 180
Db 226 VTGLVDLPSLG 237

RESULT 8
Q8N4C7 PRELIMINARY; PRT; 294 AA.
AC Q8N4C7
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Similar to RIKEN CDNA A030009B12 gene.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Strausberg R.;
RC Tissue=Skin;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034696; AAH34696.1; -.
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0008565; P:protein transporter activity; IEA.
DR GO; GO:0006886; P:intracellular protein transport; IEA.
DR InterPro; IPR006012; Syntaxin.
DR InterPro; IPR000727; T_SNARE.
DR Pfam; PF05739; SNARE; I.
DR SMART; SM00397; t_SNARE; 1.
DR SMART; PF00804; Syntaxin; 1.
DR PROSITE; PS00914; SYNTAXIN; 1.
DR PROSITE; PS0192; T_SNARE; 1.
SQ SEQUENCE 294 AA; 34324 MW; 1A7E96916DCE57 CRC64;

Query Match
Best Local Similarity 23.8%; Score 96; DB 4; Length 294;
Matches 44; Conservative 31; Mismatches 58; Indels 52; Gaps 10;

QY 27 EEINKAIDDAIAAEQSETI-----PDHAKFERHGVIVDF-----KGE- 70
Db 103 EYINRSLNDLVKKEKSEVNGPSSVVTIRLKSQAAMP-RHFQQIMFYNDTIAAKQEK 161
QY 71 ---LAVNTEARG-----LKQKQGDANYKGBEG-----GIVKAHL---LIGVHDDIVS 113
Db 162 CKTFIDRLQEVAKENSEEDVDMLHQKWEVFNESLLTEINITKAQSEISQRKELVN 221
QY 114 MEYDLAYKGLDHPHTHVISDIQDFVALSLEISDEG-----NITMTSEVRFQFANVNH 169
Db 222 LE-----NQIKLRDLFIQISLLVEQGESINNIENTVNSTKEVYVNTKEK 267
```

```
QY 170 GGLSI 174
Db 268 FGLAV 272

RESULT 9
Q81WAO PRELIMINARY; PRT; 959 AA.
AC Q81WAO
DT 01-JUN-2003 (TREMblrel. 24, Created)
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)
DE Prophage Lambda02, tape measure protein, putative.
GN BA4082.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
SEQUENCE FROM N.A.
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rillstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman J.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.B., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86 (2003).
DR EMBL; AE017036; AAP27807.1; -.
DR TIGR; BA4082; -.
KW Complete proteome.
SQ SEQUENCE 959 AA; 104502 MW; F70773DD6A500055 CRC64;

Query Match
Best Local Similarity 23.5%; Score 96; DB 16; Length 959;
Matches 51; Conservative 33; Mismatches 57; Indels 76; Gaps 11;

QY 23 DKITEINKAIDDAIAAEQSETI-----DPMKVP-DHAKFERHGVIVDFKGE- 70
Db 764 NVIESINKEIEVA---TQAEKLEEGIKRYQGDITLMPDFQAQKFEQALQVADKNVQQ 819
QY 71 -----LAMRNIEARGLQKQGDANYKGBEGIVKAHLIGVHDDIVSME----- 115
Db 820 FYTKAKETSISKEIEAGGMLSL---DAGKRFESIIVY---EDGVKSLDKQTKGW 870
QY 116 ---YDLAYKGLDHPHTHVISDIQDFVALSLEISDEGNTMTSEVRFQFANVNHIGGL 172
Db 871 RENYERAFKLGEIKP-----EERKATLDAIALYE-----SKHVNDL 906
QY 173 -SILDPIFGVLSVDLTAIFQDTVRKEMTKVLAPAFKR 208
Db 907 QSIENDGFKVL-----QOHMKEDAEVLAQAKR 935

RESULT 10
Q8KVC7 PRELIMINARY; PRT; 186 AA.
AC Q8KVC7
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
```

RESULT 13
Q9V1N1

Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R., Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S., Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M., Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D., Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G., Fraser C.M.;

RT "Complete genome sequence and comparative genomic analysis of an RT emerging human pathogen, serotype V *Streptococcus agalactiae*.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).

DR EMBL; AE014265; AAN00484.1; --

DR TIGR; SAG1620; --

DR InterPro; IPR005289; GTP-binding_dom.

DR InterPro; IPR006073; GTP_OBG.

DR InterPro; IPR002917; MMR_HSRL.

DR InterPro; IPR005225; Small_GTP.

DR Pfam; PF01926; MMR_HSRL; 1.

DR PRINTS; PR00326; GTP1OGB.

DR TIGRFAMs; TIGR00650; MG442; 2.

DR TIGRFAMs; TIGR00231; small_GTP; 2.

KM Complete proteome.

SQ SEQUENCE 436 AA; 48981 MW; E2C064B3F04B8644 CRC64;

Query Match 8.7%; Score 93; DB 16; Length 436;

Best Local Similarity 22.2%; Pred.No. 22;

Matches 47; Conservative 39; Mismatches 82; Indels 44; Gaps 10;

QY 30 NKAIDDAIAAIEOSEIDPMKVPDHDADKPERHVGIVDFKGLAMRNIEAGLKQMKRQGD 89

DB 21 NRIAGERISIVEDVEGVTDRDRIYTTGWLNRKFSLIDTGG--IDVDAPFMQEIKHQAD 77

QY 90 AN-----VKGEIGIVKAAH-----LLIGVHD-DIVSVMEYDJ-----AYK 121

DB 78 IAMTEADVIVFVSGKEGVDADEYVSRILYKTKPVILAVKNKVDPEVNDIVDFVSLG 137

QY 122 LGLDHPPTHV-----ISDIQDFVVALSLEISDEG-NITMTSFVFRQFANV-----VNHIGG 171

DB 138 LGGDYPFLSSVHGHTGTDILDAIVE-NLPVEEENENPDIIRFSLIGRPNVGKSLINAILG 196

QY 172 LS-ILDPFIFGVLSDVLTAFQDVTVRKEMTKV 201

DB 197 EDRIASPVAGTTRDAIDTNFVDSQCGEYTM 228

RESULT 15

ID Q9XIJ5 PRELIMINARY; PRT; 697 AA.

AC Q9XIJ5;

DT 01-NOV-1999 (T-EMBLrel. 12, Created)

DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE T10024.19.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eucoids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI_TaxId=3702;

RP [1]

RP SEQUENCE FROM N.A.

RA Shinn P., Brooks S., Buehler E., Chao Q., Dunn P., Khan S., Kim C., Walker M., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez C., Hansen N.F., Huizar L., Krenetskaia I., Lenz C., Li J., Liu S., Luros S., Rowley D., Schwartz J., Tortum M., Vysotskaya V., Yu G., Davis R.W., Federpiel N.A., Theologis A., Ecker J.R.;

RA "Genomic sequence for Arabidopsis thaliana BAC T10024 from Chromosome 1.";

RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL; AC007067; AAD39579.1; --

DR PIR; B86239; B86239.

DR InterPro; IPR008938; ARM.

DR InterPro; IPR000225; Armadillo.

DR InterPro; IPR003613; Znf_mogRING.

DR InterPro; IPR001841; Znf_ring.

DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; cpn60_TCP1; 1.
 DR ATP-binding; Chaperone.
 FT NON TER 1 186 186
 SQ SEQUENCE 186 AA; 20156 MW; 4114C4B4FAB3B600 CRC64;
 Query Match 8.5%; Score 90.5; DB 2; Length 186;
 Best Local Similarity 24.2%; Pred. No. 11;
 Matches 30; Conservative 23; Mismatches 44; Indels 27; Gaps 5;
 QY 7 IAAFAFVAVSADPIHYDKITEINKAIDDAIAAI-----EQSETIDPMKVPDHDADKFER- 60
 DB 3 VLAQALIGVLGNVTADANPMDLKGIDKXAKVVENIAEQAEV-----GDKFEKI 54
 QY 61 -HVGIVDFKGLAMRNIEARGLKQMKRQGDANV---KGEIGVKAHLIGVHDDIVSMVEY 116
 DB 55 EHVAKISANGDETTIGKLIAEAMGKVKKEGVITVEAKGTETTV-----DVVEGMOF 105
 QY 117 DLAY 120
 DB 106 DRGY 109
 RESULT 25
 Q18126 PRELIMINARY; PRT; 681 AA.
 ID Q18126 AC
 AC Q18126 AC
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN C24A3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 R None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC STRAIN=Bristol N2;
 RX Zavallo A.;
 RT "The sequence of C. elegans cosmid C24A3.";
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC Waterston R.;
 RA "Direct Submission.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U40424; AAA91461.1; -.
 DR FIR; T15590; T15590.
 DR WormPep; C24A3.1; CE04058.
 DR InterPro; IPR008627; GETHR.
 DR Pfam; PF05671; GETHR; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 681 AA; 79533 MW; 2DA37C9046C26A20 CRC64;
 Query Match 8.5%; Score 90.5; DB 5; Length 681;
 Best Local Similarity 25.1%; Pred. No. 63;
 Matches 51; Conservative 41; Mismatches 62; Indels 49; Gaps 11;
 QY 24 KITEINKAIDD---AIAAIE-----QSETI-----DPMKVPDHDADKFERHVGIVDFKGL 71
 DB 133 KMTENQOKAKDDFNGRISVLESMTIRAQSERIVANENMR-----HSPFAKLAELSGLLEL 187

DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; cpn60_TCP1; 1.
 DR ATP-binding; Chaperone.
 FT NON TER 1 186 186
 SQ SEQUENCE 186 AA; 20156 MW; A114C4B4FAB3B600 CRC64;
 Query Match 8.5%; Score 90.5; DB 2; Length 186;
 Best Local Similarity 24.2%; Pred. No. 11;
 Matches 30; Conservative 23; Mismatches 44; Indels 27; Gaps 5;
 QY 7 IAAFAFVAVSADPIHYDKITEINKAIDDAIAAI-----EQSETIDPMKVPDHDADKFER- 60
 DB 3 VLAQALIGVLGNVTADANPMDLKGIDKXAKVVENIAEQAEV-----GDKFEKI 54
 QY 61 -HVGIVDFKGLAMRNIEARGLKQMKRQGDANV---KGEIGVKAHLIGVHDDIVSMVEY 116
 DB 55 EHVAKISANGDETIGKLIAEAMGVKKEGVITVEAKGTETTV-----DVVEGMOF 105
 QY 117 DLAY 120
 DB 106 DRGY 109
 RESULT 25
 Q18126 PRELIMINARY; PRT; 681 AA.
 ID Q18126 AC
 AC Q18126
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN C24A3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Telodermata; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 R None;
 RT 'Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.';
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX Pavallo A.;
 RA 'The sequence of C. elegans cosmid C24A3.';
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RA "Direct Submission.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U40424; AAA91461.1; -.
 DR FIR; T15590; T15590.
 DR WormPep; C24A3.1; CE04058.
 DR InterPro; IPR008627; GETHR.
 DR Pfam; PF05671; GETHR; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 681 AA; 79533 MW; 2DA37C9046C26A20 CRC64;
 Query Match 8.5%; Score 90.5; DB 5; Length 681;
 Best Local Similarity 25.1%; Pred. No. 63;
 Matches 51; Conservative 41; Mismatches 62; Indels 49; Gaps 11;
 QY 24 KITEINKAIDD---AIAAIE-----QSETI-----DPMKVPDHDADKFERHVGIVDFKGL 71
 DB 133 KMTENQOKAKDDFNGRISVLESMTIRAQSERIVANENMR-----HSPFAKLAELSGLLEL 187

RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437065; AAM49326.1; -;
DR GO; GO:0005524; F:ATP binding; IEA.
DR InterPro; IPR002423; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cn60/TCP-1.
DR Pfam; PF00118; Cn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON TER 1 1
SQ SEQUENCE 186 AA; 20095 MW; 51084F7D1395090A CRC64;

Query Match 8.4%; Score 90; DB 2; Length 186;
Best Local Similarity 25.4%; Pred. No. 13;
Matches 30; Conservative 23; Mismatches 31; Indels 34; Gaps 6;
QY 13 VAVSADTHYDKITEEINKAIDDAIAI-----EQSETIDPMKVPDHDKFER--HVGIV 65
DB 16 VTAGANPM-----DLKRGIDKAVAKVKNIAQAEEV-----GDKFEXIEHVAKI 60
QY 66 DFKELANRNTAARGLKMKGQGANV---KGEIGIVKAHLIGVHDDIVSMYDILAY 120
DB 61 SANGDEAIGKLIAGAMQVKEGIVTVEEAKGTETV-----DVVEGMQFDRGY 109

RESULT 28
Q8DD69 PRELIMINARY; PRT; 333 AA.
AC Q8DD69;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glyceraldehyde-3-phosphate dehydrogenase.
GN Vv11141.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RT Choy H.E.;
RL "Complete genome sequence of *Vibrio vulnificus* CMCP6.";
DR EMBL; AE016800; AAO09615.1; -;
DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR006424; GAPDH-I.
DR InterPro; IPR000173; GAP-chohydrogenase.
DR Pfam; PF00044; gpdh; 1.
DR Pfam; PF02800; gpdh_C; 1.
DR TIGRfams; TIGR01534; GAPDH-I; 1.
DR PROSITE; PS00071; GAPDH; 1.
KW Complete proteome.
SQ SEQUENCE 333 AA; 36634 MW; 87E1D9698395F681 CRC64;

Query Match 8.4%; Score 90; DB 16; Length 333;
Best Local Similarity 24.2%; Pred. No. 27;
Matches 48; Conservative 35; Mismatches 63; Indels 52; Gaps 9;
QY 59 EHHVGIYDFPKG-----ELAMENIARGLKMKRGK-----DANVKGEGIVKAHLIGV 107
DB 78 QRDASIDVSGCDVVIATGVVHRKSRFLNQVLEGGVVRVVSAPVK--EEGI--ANIVGV 134
QY 108 HDDIVSMYD-----LAYKLG-----DLHPTTHVISDQDFVVA---- 141
DB 135 NDEIFNPQHRIVTAASCTTNCIAPVVKVIHEKLGIEQASFTTHDUTNTQITLDAPHKD 194

QY 142 -----LSLEISDEGNITMTSPFVROFANVNVHIGLSILDPIFGV-LSDVLTAFQD 192
DB 195 LRARACGSLIPTTGTSAI--VEFPPELKORINGHAVRVPLANASLTDIIFDVKRD 251
QY 193 TVRKEMTKVLAPAFKREL 210
DB 252 TTAEEVNALKEASQDEL 269

RESULT 29
O96275 PRELIMINARY; PRT; 1558 AA.
AC O96275;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RESA-H3 antigen.
DE PF0015W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalloom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*.";
RT Science 282:1126-1132(1998).
RL EMBL; AB001424; AAC71972.1; -;
DR PIR; B71603; B71603.
SQ SEQUENCE 1558 AA; 175658 MW; E2FC68E9036C81BC CRC64;

Query Match 8.4%; Score 90; DB 5; Length 1558;
Best Local Similarity 23.3%; Pred. No. 2.1e+02;
Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;
QY 22 YDKITEEINKAIDDAIAAEQSETIDPMKVPDHDKFERHVGIVDFKGLAMRNIARGL 81
DB 747 FNTVLDKVEETVEISGESLENNE---MDKAPFFSEIFDNVKGIGENLLTGMFRSITSIV 802
QY 82 KMKQEGDANVKGEGIVKAHLIGVHDDIVSMYDILAYKGLDHPHPTTHVISDQDFVVA 141
DB 803 IQSEKVDLN---ENVVSSIL-----DNIEKGLLNKLENISSTEGVQETVTEHV-- 851
QY 142 LSLEISDEGNITMTSPFV---RQFANVNVHIGLS-----ILDPIFGVLSVLTAF--IFQ 191
DB 852 -----EQNV-YVDVDVPAMKQDFLGILNEAGGLKEMFFNLEDVFKSESVDITVEIKD 903
QY 192 DTVRKEMTKVLAPAFKRELKN 213
DB 904 EPVQKEVEKETVSIIE-EEMEN 924

RESULT 30
Q9U0P0 PRELIMINARY; PRT; 1786 AA.
AC Q9U0P0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Liver stage antigen-3 precursor.
GN LSA-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;

Query Match 8.4%; Score 90; DB 5; Length 1558;
Best Local Similarity 23.3%; Pred. No. 2.1e+02;
Matches 47; Conservative 43; Mismatches 78; Indels 34; Gaps 9;
QY 22 YDKITEEINKAIDDAIAAEQSETIDPMKVPDHDKFERHVGIVDFKGLAMRNIARGL 81
DB 747 FNTVLDKVEETVEISGESLENNE---MDKAPFFSEIFDNVKGIGENLLTGMFRSITSIV 802
QY 82 KMKQEGDANVKGEGIVKAHLIGVHDDIVSMYDILAYKGLDHPHPTTHVISDQDFVVA 141
DB 803 IQSEKVDLN---ENVVSSIL-----DNIEKGLLNKLENISSTEGVQETVTEHV-- 851
QY 142 LSLEISDEGNITMTSPFV---RQFANVNVHIGLS-----ILDPIFGVLSVLTAF--IFQ 191
DB 852 -----EQNV-YVDVDVPAMKQDFLGILNEAGGLKEMFFNLEDVFKSESVDITVEIKD 903
QY 192 DTVRKEMTKVLAPAFKRELKN 213
DB 904 EPVQKEVEKETVSIIE-EEMEN 924

RESULT 30
Q9U0P0 PRELIMINARY; PRT; 1786 AA.
AC Q9U0P0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Liver stage antigen-3 precursor.
GN LSA-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;

Query Match 8.4%; Score 90; DB 16; Length 333;
Best Local Similarity 24.2%; Pred. No. 27;
Matches 48; Conservative 35; Mismatches 63; Indels 52; Gaps 9;
QY 59 EHHVGIYDFPKG-----ELAMENIARGLKMKRGK-----DANVKGEGIVKAHLIGV 107
DB 78 QRDASIDVSGCDVVIATGVVHRKSRFLNQVLEGGVVRVVSAPVK--EEGI--ANIVGV 134
QY 108 HDDIVSMYD-----LAYKLG-----DLHPTTHVISDQDFVVA---- 141
DB 135 NDEIFNPQHRIVTAASCTTNCIAPVVKVIHEKLGIEQASFTTHDUTNTQITLDAPHKD 194

Query Match 8.4%; Score 90; DB 16; Length 333;
Best Local Similarity 24.2%; Pred. No. 27;
Matches 48; Conservative 35; Mismatches 63; Indels 52; Gaps 9;
QY 59 EHHVGIYDFPKG-----ELAMENIARGLKMKRGK-----DANVKGEGIVKAHLIGV 107
DB 78 QRDASIDVSGCDVVIATGVVHRKSRFLNQVLEGGVVRVVSAPVK--EEGI--ANIVGV 134
QY 108 HDDIVSMYD-----LAYKLG-----DLHPTTHVISDQDFVVA---- 141
DB 135 NDEIFNPQHRIVTAASCTTNCIAPVVKVIHEKLGIEQASFTTHDUTNTQITLDAPHKD 194

Qy 61 -HVGI VDFK GELAMRNIEARG LKQMKRQGDANV--KGE GIVKAHLLIGVHDDIVSMEY 116

244 SRIENHPAETI.SNVDQATEDTRNMDSLREKSAQIHTANKEATESNRLKSOFLANISHE- 302

61 -HVGTVDKGEI.AMPNTEAPGLKOMKPOGDANY--KCEEGTIVKAHL.I.GVHDDIVSMEY 116

```

QY 97 GIVKAHL-LIGVHDDIVSMYDLAYKGLDHLPTT-----HVISDIQDF-----VVA 141
D 303 --IRTPLNAIMGY---IQLKQDQLDQORLYMDTIAQSTNNLITLTIISDILDSKIEAGK 357
QY 142 LSLRISDEG--NITMTSFVRQFANVNHIGGILDPFGVLSDLVLTAFQDTV--RKE 197
D 358 LSLDLTDINIRIDVEYQILSSANLVASKELDLI--PEFDL--DVTWIGDPVIRQV 414
QY 198 MTKVLAPAFK 207
D 415 LTNLIGNAIK 424

RESULT 33
Q7WNV1 PRELIMINARY; PRT; 279 AA.
AC Q7WNV1;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE Putative exported protein.
GN BB0934.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640439; CAE1433.1; -
KW Complete proteome.
SQ SEQUENCE 279 AA; 30292 MW; 0092B34E2CD15E26 CRC64;

Query Match 8.3%; Score 88.5; DB 16; Length 279;
Best Local Similarity 24.4%; Pred. No. 28;
Matches 53; Conservative 33; Mismatches 74; Indels 57; Gaps 12;

QY 3 KFLIIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-ATAAIEQSETID---PMK 50
D 30 RYLATAVVALAVAAALYARQDPEHFDEV---VLQCLDQAMSTVRRGSDGILPAC 86
QY 51 VPDHADFEHVGIVDFKGLAMRNIEAR-GLKQMKQGDANVKGEGIVKAHL-----104
D 87 QPELAELVGEDFGAVD---EPTLRVLATVAANFAEYSSSAVAYEAIEKSOHLNCGNL 143
QY 105 -----IGVHDDIVSMYDLAYKGLD-----LHPTTHVISDIQDFVAL 142
D 144 VFLMGVLYGLGKKEGIRPIGVGGAVGNAHQIMYKSGNNRLLDPTTGVVAAL-TFNKLM 202
QY 143 SLEISDEGNI-----TWTSFVRQFANVNHIGG 171
D 203 SGQVPPEQVFRVFRVKAKTIDSFRDRVYRAVG--GG 237

Query Match 8.3%; Score 88.5; DB 16; Length 279;
Best Local Similarity 24.4%; Pred. No. 28;
Matches 53; Conservative 33; Mismatches 74; Indels 57; Gaps 12;

QY 3 KFLIIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-ATAAIEQSETID---PMK 50
D 30 RYLATAVVALAVAAALYARQDPEHFDEV---VLQCLDQAMSTVRRGSDGILPAC 86
QY 51 VPDHADFEHVGIVDFKGLAMRNIEAR-GLKQMKQGDANVKGEGIVKAHL-----104
D 87 QPELAELVGEDFGAVD---EPTLRVLATVAANFAEYSSSAVAYEAIEKSOHLNCGNL 143
QY 105 -----IGVHDDIVSMYDLAYKGLD-----LHPTTHVISDIQDFVAL 142
D 144 VFLMGVLYGLGKKEGIRPIGVGGAVGNAHQIMYKSGNNRLLDPTTGVVAAL-TFNKLM 202
QY 143 SLEISDEGNI-----TWTSFVRQFANVNHIGG 171
D 203 SGQVPPEQVFRVFRVKAKTIDSFRDRVYRAVG--GG 237

RESULT 34
Q7W160 PRELIMINARY; PRT; 279 AA.
ID Q7W160
AC Q7W160;
DT 01-OCT-2003 (TRENBLrel. 25, Created)

```

```

DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative exported protein.
GN BP0840.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640425; CAE40249.1; -
KW Complete proteome.
SQ SEQUENCE 279 AA; 30292 MW; 0092B34E2CD15E26 CRC64;

Query Match 8.3%; Score 88.5; DB 16; Length 279;
Best Local Similarity 24.4%; Pred. No. 28;
Matches 53; Conservative 33; Mismatches 74; Indels 57; Gaps 12;

QY 3 KFLIIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-ATAAIEQSETID---PMK 50
D 30 RYLATAVVALAVAAALYARQDPEHFDEV---VLQCLDQAMSTVRRGSDGILPAC 86
QY 51 VPDHADFEHVGIVDFKGLAMRNIEAR-GLKQMKQGDANVKGEGIVKAHL-----104
D 87 QPELAELVGEDFGAVD---EPTLRVLATVAANFAEYSSSAVAYEAIEKSOHLNCGNL 143
QY 105 -----IGVHDDIVSMYDLAYKGLD-----LHPTTHVISDIQDFVAL 142
D 144 VFLMGVLYGLGKKEGIRPIGVGGAVGNAHQIMYKSGNNRLLDPTTGVVAAL-TFNKLM 202
QY 143 SLEISDEGNI-----TWTSFVRQFANVNHIGG 171
D 203 SGQVPPEQVFRVFRVKAKTIDSFRDRVYRAVG--GG 237

RESULT 35
Q822J2 PRELIMINARY; PRT; 1393 AA.
ID Q822J2
AC Q822J2;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE DNA-directed RNA polymerase, beta' subunit.
GN RPOC OR CCA00690.
OS Chlamydomonadales; Chlamydiales; Chlamydiales; Chlamydiales;
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales;
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzaple E., Khouri H., Federova N.B., Carty H.A.,
RA Umavam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomonadales caviae (Chlamydia psittaci GPIC):
examining the role of niche-specific genes in the evolution of the

```

Weber N., Grueninger D., Schmidheini T., Mewes H.W., Lemcke K., Mayer K.F.X.;
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR ENBL; ALU035528; CAB36839.1; -
 DR ENBL; AL161537; CAB78417.1; -
 DR PIR; T05244; T05244.
 DR Hypothetical protein.
 SW Hypothetical protein.
 KO SEQUENCE 2137 AA; 240107 MW; 0A3AFCB275C2BCB2 CRC64;
 Query Match 8.3%; Score 88.5; DB 10; Length 2137;
 Best Local Similarity 21.8%; Pred. NO. 4.2e+02;
 Matches 45; Conservative 34; Mismatches 82; Indels 45; Gaps 10
 QY 35 DAIAAEQSEITDPMKVPDHADK----FERHVGIVDFKGLAM--RNIEAR-GLKQMRQ 87
 DB 614 DSHFILELVQNAADNDKYPEHEVPELTILQKTGIIVLNNECGFMPENIRALCDVGQSTKK 673
 QY 88 GDANKVGEIGI-VKXHLIG-----VHDDIVSMYDIAY-KLGDLIPTT---HVISDIQDF 138
 DB 674 GSGGVIGKGGKGFKXSVRVSDAPEIHSGHFHFKFDISEQGIYLPVTVPVPHDISLSM 733
 QY 139 VVALSLEISDSG---NTMTSFVRQFANVNHIGLSILDRPF----- 179
 DB 734 LSGRAHLKDGAWTCITILFRAIDSRITVNI-----EFMSDLHPSLLLFLRLQC 787
 QY 180 ----GVLSDVITAIQDTPVRKEMTKV 201
 DB 788 IYVRNVLDSDSLVWVRKEVWSKNIVKV 813
 RESULT 37
 Q8FGF6 PRELIMINARY; PRT; 609 AA.
 ID Q8FGF6
 AC Q8FGF6;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DE 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Putative ABC transporter protein.
 GN C2421.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06:HI / CFT073 / ATCC 700928;
 RC MEDLINE=22388234; PubMed=12471157;
 RX Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 RL ENBL; AB016762; AAN80880.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
 DR GO; GO:0006106; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC TM transpt.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00654; ABC membrane; 1.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER 1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER 2; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER 2; 1.

```
KW Complete proteome.
SQ SEQUENCE 609 AA; 67758 MW; F7FB7AFF75E07378 CRC64;

Query Match      8.2%; Score 88; DB 16; Length 609;
Best Local Similarity 22.2%; Pred. No. 88;
Matches 43; Conservative 26; Mismatches 65; Indels 60; Gaps 6;

QY 6 LIAAVAFVAVSADP-----IHYKITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERH 61
DB LIAAIVAMIRFAEPMAFISYTSVVE-----LIASALQRIEQFMAIAPLPVAEQSEMPERY 351
QY 62 VGIIVDFKXGELAMRNIEARGLKQMKRQGDANVKGBEGIVKAHLIGVHDDIVSMEYDLAYK 121
DB 352 -----DIRFDNVSFR-----YE 363

QY 122 LGDLHPTHTVSDIQDFVVALSLEISDEGNITMTSFVRQFANVNVN---HIGGLSILDP 178
DB 364 EGDGHALNHVSLTPPAASMSALVCAGSAGKTTVTKLMR-YADPQQQGISIGGVDIRRLT 422

QY 179 FGVLSDLVLTAFQD 192
DB 423 PEQLNSLISVVFQD 436

RESULT 38
Q897K6 PRELIMINARY; PRT; 485 AA.
AC Q897K6;
DT 01-JUN-2003 (TREMREL. 24, Created)
DT 01-JUN-2003 (TREMREL. 24, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Uroporphyrinogen III synthase/methyltransferase (EC 4.2.1.75).
CN CTC00728
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baeumer S., Fricke W.F., Wierze A., Liesegang H.,
RA Becker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RA "The genome sequence of Clostridium tetani, the causative agent of
RA tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR EMBL; AE015938; AA055330.1;
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008168; F:methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0004852; F:uroporphyrinogen-III synthase activity; IEA.
DR GO; GO:0006783; P:heme biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000878; Cor/por_Mettransf.
DR InterPro; IPR003754; HEM4 synth.
DR InterPro; IPR003043; Uropor_Mettransf.
DR Pfam; PF02602; HEM4; 1.
DR Pfam; PF00590; TP_methylase; 1.
DR PROSITE; PS00839; SUMT_1; 1.
KW Lysase; Methyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 485 AA; 54624 MW; B09D45F85570F747 CRC64;

Query Match      8.2%; Score 87.5; DB 16; Length 485;
Best Local Similarity 20.9%; Pred. No. 71;
Matches 44; Conservative 51; Mismatches 83; Indels 33; Gaps 10;

QY 27 BEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHV-GIVDFKXGELAMRNIEARGL---K 82
DB 262 KELREKSLDLGAQVTE---INAIKSRDDELEKYVERLKYDK-FIVITSVNAVNLFDY 317
QY 83 QMKRQGD-ANVKGBEGIV-----KAHLIGVHDDIVSMEYDLAYKGLDHPHTHTVSDIQ 136
DB 317 -----K 136
```

```
DB 318 LVKKEYDIRNKGKGFANVIGPATEKALKKRGKADIIAEKFEVADLFNKLKP--VINKDDK 375
QY 137 DFV-----VALSLEISDEGNITMTSFVRQFANVNVNHHIGGLSILDPFIF-----GVL 182
DB 376 IFFPHSKQARQVLYEALKDKG---CHVEEVYSYEVWKGVEVNFLEIFDDVDIVLTSPSIV 432
QY 183 SDVLTAFQDVTURKMTKVLAPAFKRELEKN 213
DB 433 RNMDIMGLEKLEKKTTSISIGPITAKELDKN 463

RESULT 39
Q8IK83 PRELIMINARY; PRT; 1620 AA.
AC Q8IK83;
DT 01-MAR-2003 (TREMREL. 23, Created)
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Hypothetical protein.
CN PF14_0723.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368664;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mangall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RA "Genome sequence of the human malaria parasite Plasmodium
RA falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014828; AA037336.1;
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR002181; Fibrinogen C.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR004043; LCCL_dom.
DR InterPro; IPR008997; RicinB like.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF03815; LCCL; 1.
DR SMART; SM00603; LCCL; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS50022; FA58C_3; 1.
DR PROSITE; PS50820; LCCL; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 1620 AA; 184874 MW; 954B08097161F049 CRC64;

Query Match      8.2%; Score 87.5; DB 5; Length 1620;
Best Local Similarity 22.9%; Pred. No. 3.5e+02;
Matches 43; Conservative 37; Mismatches 53; Indels 55; Gaps 9;

QY 61 HVGIVDFKXGELAMRNIEARGLKQMKRQGDANV-----KGBEGIVKAHLIGVHDDIVSM 114
DB 810 HSDIIDNKGGLVNTIES-GMDHYVGSINNINIESISLNKQKGLL-----DIPE 858
QY 115 EYDLAYKGD-----LHPTHTVSDIQDFVVAL-----SLSD 148
DB 859 EKEGNTNIRESSIFHHKTVSSLIEDCPDLJLFLNQTSTFLEKGNIRNNKGTLYKND 918
QY 149 EGNITMTSFVRQFANVNVNHHIGGLSILDPFVGLSDVLTAFQDVTUR--KEMTKVLAPA- 205
```



```
Db 919 DENWT-----VKNFHELISNL--MENIDAHGVDSVISVQETIRIIEKTKKLPAD 971
QY 206 --FKRELE 211
Db 972 MLGKKQIE 979

RESULT 40
Q8KV75 PRELIMINARY; PRT; 186 AA.
AC Q8KV75;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN Cpn60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.I., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
RT throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437156; F:ATP binding; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20155 MW; E78884004BACABED CRC64;

Query Match 8.1%; Score 87; DB 2; Length 186;
Best Local Similarity 24.6%; Pred. No. 22;
Matches 29; Conservative 23; Mismatches 32; Indels 34; Gaps 6;

QY 13 VASADPIHVDKITEINKAIDAI--EQTIDPMKVPDHPADKFER--HVGIV 65
Db 16 VTAGAPM-----DKRTDRAVAKVVENIAEQAEV-----GDKFEKIEHVAKI 60

QY 66 DFKGELAMRNIEARGLKQMKRQGDANY---KGEGIVKXHLIGVHDDIVSMYDILAY 120
Db 61 SANGDEAIGKLIARQVKEGVIIVBEAKGTETV-----DVVEGMCQFDRGY 109

RESULT 41
Q9ZVB0 PRELIMINARY; PRT; 372 AA.
AC Q9ZVB0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F9K20.3.
GN F9K20.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,
RA Kremenetskaia I., Lueros J., Araujo R., Buehler E., Conway A.B.,
RA Dewar K., Feng J., Kim C., Li Y., Shinn P., Davis R.W., Ecker J.R.,
RA Federapfel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F9K20 sequence.";
RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005679; AAC83019.1; -.
DR PIR; G96818; G96818.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR SMART; SM00028; TPR; 4.
KW Hypothetical protein.
SQ SEQUENCE 372 AA; 41053 MW; AFD052A81D1877A9 CRC64;

Query Match 8.1%; Score 87; DB 10; Length 372;
Best Local Similarity 24.8%; Pred. No. 55;
Matches 40; Conservative 27; Mismatches 58; Indels 36; Gaps 8;

QY 27 EINKAIDDAIAIEQSETIDPMKVPDHPADKFERHVG---IVDPKGELAMRNIEARGLKQ 83
Db 38 EEKDPALQQRKSSKQSVSV-PRKAPGLNTQFEKSGRSFDIDFDERL--ENIRRSALQ 94

QY 84 MKRQ-----GDANVKGEGIVKXHLIGVHDDIVSMYDILAYKLGDLHPT----- 128
Db 95 KTEVVEKFGFIDYDAPVKSQKTIGLTKGV--GIAVVVGLVFGALGDFLPTGSDSPT 152

QY 129 --THVISDIQDFVVALSLEISDEGNITMTSFEVRQFANVN 167
Db 153 KNTTVVKN-----QISEEEKATLQ--RLKEFETLN 182

RESULT 42
Q8GWV1 PRELIMINARY; PRT; 385 AA.
AC Q8GWV1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Atlg78915).
GN Atlg78915.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RT Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK118621; BAC43219.1; -.
DR EMBL; BT008318; AAP37677.1; -.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR SMART; SM00028; TPR; 4.
KW Hypothetical protein.
SQ SEQUENCE 385 AA; 42579 MW; 6608867205554E48 CRC64;

Query Match 8.1%; Score 87; DB 10; Length 385;
Best Local Similarity 24.8%; Pred. No. 58;
Matches 40; Conservative 27; Mismatches 58; Indels 36; Gaps 8;
```

```
QY 27 EEINKAIDDAIAEIOSETIDPMKVPDADKFEHVG---IVDFKGLAMNIEARGLQ 83
Db 38 EEKDPALQQRSSKQSVV-FRKAGLNTQFEKSGRSFDIDFDEL--ENIRSALEQ 94
QY 84 MKRQ-----GDANKGEEGIVKAHLIGVHDDIVSMYDIAVYKLGDLHPT-----128
Db 95 KKEVXKEFGIDYDAFKSDQKTLIGTKGVV--GIAVVVGLVFLGDLFLTGSDSPT 152
QY 129 --THVDSIQDFVVALSLEISDEGNITMTSFEVRQFANVN 167
Db 153 KNTTWKXN-----QISEEKATLQO-RLKEFEFTLLN 182

RESULT 43
ID O44999 PRELIMINARY; PRT; 399 AA.
AC O44999;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN W03D8.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.,"
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Jones K., Graves T., Ozersky P.;
RT "The sequence of C. elegans cosmid W03D8.,"
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043702; AAK21491.1; -.
DR PIR; T32933; T32933.
DR WormPep; W03D8.5; CE18325.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000535; MSP domain.
DR InterPro; IPR008962; PapD-like.
DR Pfam; PF00635; MSP domain; 1.
DR PROSITE; PS02020; MSP; 1.
KW Hypothetical protein.
SQ SEQUENCE 399 AA; 45411 MW; 050FE0780CB62190 CRC64;

Query Match 8.1%; Score 87; DB 5; Length 399;
Best Local Similarity 20.3%; Pred. No. 61;
Matches 40; Conservative 38; Mismatches 63; Indels 56; Gaps 7;

QY 11 AFVAYSADPIHYD-----KITEINKAIDDAIAEIOSETIDPMKV 51
Db 44 SLVVVAAPVQDFRFAENQETQTKDSEEMGEVEEKKKKKEENKAVEKVEPKE-102
QY 52 PDHAKPFHVGIVDFKGLAMNIEARGLQKMRQGDANVKGEGIVKAHLIGVHDDI 111
Db 103 ----EKIEKKLEKMSKKEEKEEVEPKEEKKKEEKGVNESRK-----IPDTN 153
QY 112 VSMYDIAVYKLGDLHPTTH-----VISDIQDFVVALSLEISDEGNITMTSFEVRQFANV 165
Db 154 TGKSHVSAVPPQLLFTPDCEQEKLTITNTHDKKIMYKGLSD--NI---AFQ-----201
```

```
QY 166 VNHIGLSILDPIFGVL 182
Db 202 -----LNPVFGVL 209
QY 202 -----LNPVFGVL 209

RESULT 44
ID Q8U2M6 PRELIMINARY; PRT; 562 AA.
AC Q8U2M6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein PF0807.
GN PF0807.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.,"
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010197; AAL80931.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 562 AA; 64247 MW; 16EA3FE3946BB0FA CRC64;

Query Match 8.1%; Score 87; DB 17; Length 562;
Best Local Similarity 22.7%; Pred. No. 95;
Matches 51; Conservative 36; Mismatches 70; Indels 68; Gaps 12;

QY 22 YDKITEINKAIDDAIAEIOSETIDPMKVPDADKFEHVG-IVD-----FKG-----69
Db 63 YELTSLINFLIE-----EQNERLNFLLP--FLEFHRGIGKIEGQRAFFKGMMLSGV 114
QY 70 -----ELAMRNIARGLKQKROGDANV-KGEEGIVKAHLIGVHDDIVSMYDYL 118
Db 115 ANMKEGISESELALSNIESLEFTEDEGKRLNVDKLERVELVSAYEFLARKVNA 174
QY 119 AYKGLDHPHTHVISDIQDFVVALSLEISDEGNITMTSFEVRQFANVNHIGGLSI-----174
Db 175 -----PLVLYSDENPFYI-----NVTFYGI-ARNFSEVILHIGNESFKVKV 216
QY 175 -----LDPIF--GV-----LSDLTAIFQDVTVRKEMTKVLAP 204
Db 217 VEGFSLDYIFVSPGVYIAYAEAGNVKSNVIRINVTKIPTRIAP 261

RESULT 45
ID Q9AZS0 PRELIMINARY; PRT; 1640 AA.
AC Q9AZS0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Tail protein.
GN ORF52.
OS Bacteriophage b11286.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=151536;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21113149; PubMed=11160885;
RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RT "Analysis of six prophages in Lactococcus lactis I11403: different
genetic structure of temperate and virulent phage populations.,"
RL Nucleic Acids Res. 29:644-651(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323669; AAK08339.1; -.
```

```
SQ SEQUENCE 1640 AA; 176321 MW; CEEES9A7A47B8F18 CRC64;
Query Match 8.1%; Score 87; DB 9; Length 1640;
Best Local Similarity 20.7%; Pred. No. 3.9e+02;
Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;

QY 16 SADPHYDKITEINKAIDDAIAAEQSETIDPMKVPDHDADKFERHVGIVDFKGLAMRN 75
DB 1142 SMSKAQYDEIVKNAQKQDRTTISAACKQQT-----EVTDKAQK--THDKTVELANSKADKN 1195
QY 76 IEA-----RGLKQMKRGDANVKGEGIVK-----AHLLI----- 105
DB 1196 VKAAAEQGETVEQYTKGFKDSNLNSFDINGVNLFLHKGWNGHVSCLKGFATGTR 1255
QY 106 GVHDDIVSMEYDLAYKGLDHPHTHVIS-----DIQDFVVALSL--EISDEGNITWTS 156
DB 1256 GLAQDETALVGBEGFELAH-HPSRGIFAVGQCGPEIRNLKAGTSILPHSMSKE-FLSLT- 1312
QY 157 FEVRQFANVNHIGLSILDPIFGVLSDLVLTAFQDVTVRKMTKVLA 203
DB 1313 -----ANLPAHADGVS-----GFLSDALGWV--KSTYKDVTSVIS 1345

RESULT 46
Q9CF57 PRELIMINARY; PRT; 1640 AA.
AC Q9CF57
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Propaghe p13 protein 14.
GN P1314 OR L14388.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=111403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AEO06370; AAK05486.1; -.
DR PIR; D86798; D86798.
KW Complete proteome.
SQ SEQUENCE 1640 AA; 176320 MW; CEEES9A7A47B8F18 CRC64;

Query Match 8.1%; Score 87; DB 16; Length 1640;
Best Local Similarity 20.7%; Pred. No. 3.9e+02;
Matches 47; Conservative 45; Mismatches 73; Indels 62; Gaps 12;

QY 16 SADPHYDKITEINKAIDDAIAAEQSETIDPMKVPDHDADKFERHVGIVDFKGLAMRN 75
DB 1142 SMSKAQYDEIVKNAQKQDRTTISAACKQQT-----EVTDKAQK--THDKTVELANSKADKN 1195
QY 76 IEA-----RGLKQMKRGDANVKGEGIVK-----AHLLI----- 105
DB 1196 VKAAAEQGETVEQYTKGFKDSNLNSFDINGVNLFLHKGWNGHVSCLKGFATGTR 1255
QY 106 GVHDDIVSMEYDLAYKGLDHPHTHVIS-----DIQDFVVALSL--EISDEGNITWTS 156
DB 1256 GLAQDETALVGBEGFELAH-HPSRGIFAVGQCGPEIRNLKAGTSILPHSMSKE-FLSLT- 1312
QY 157 FEVRQFANVNHIGLSILDPIFGVLSDLVLTAFQDVTVRKMTKVLA 203
DB 1313 -----ANLPAHADGVS-----GFLSDALGWV--KSTYKDVTSVIS 1345

RESULT 47
Q7TZ25 PRELIMINARY; PRT; 280 AA.
ID Q7TZ25
AC Q7TZ25
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Conserved hypothetical protein.
GN MB1518C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RX Garnier T., Bigmelter K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dutfoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248339; CAD96185.1; -.
KW Complete proteome.
SQ SEQUENCE 280 AA; 31664 MW; 2110B7D7A51B4EE7 CRC64;

Query Match 8.1%; Score 86; DB 16; Length 280;
Best Local Similarity 28.3%; Pred. No. 46;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;

QY 33 IDDAIAAEQSETIDPMKVPDHDADKFERHVGIVDFKGLAMRNIEARGLKQMKR----- 86
DB 128 LDPVAADVADLIQATDLKVDVEPLIERYGRGRGKAAARALDLVDGGAQSPKETWLRLL 187
QY 87 -----QGDANVKGEGIVKAHLIGVHDDIVSMEYDLAYKGLDHPHTT--HVISDI 135
DB 188 LIRAGFRPQTQIAVRNENWGAHLDMGWDIKVAAEYD-----GDHLITSRVHYEKDI 242

RESULT 48
Q91432 PRELIMINARY; PRT; 303 AA.
ID Q91432
AC Q91432
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Probable transcriptional regulator.
GN PA1312.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Tolentino S., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL; AE004560; AAG04701.1; -.
DR PIR; D83482; D83482.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000847; HTH_LysR.
DR InterPro; IPR005119; LysR_subst.
DR Pfam; PF00126; HTH_1; 1.
DR Pfam; PF03466; LysR_substrate; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
```

```
KW DNA-binding; Transcription regulation; Complete proteome.
.SQ SEQUENCE 303 AA; 33062 MW; F8C201582A0BF94 CRC64;

Query Match      8.1%; Score 86; DB 16; Length 303;
Best Local Similarity 22.6%; Pred. No. 51;
Matches 28; Conservative 19; Mismatches 39; Indels 38; Gaps 4;

QY 24 KITEINKAIDDAIAAEIQSEITDPMKVPDHPADKFERHVGIVDFKGEIARNIEARGLKQ 83
Db 71 KVLDEARLAIDARA-----EHG-----ELRGLRVTITQEVGLRQ 106
QY 84 MKRQGDANVKEEGIVKAHLIGV-----HDDIVSMYDIAVKGLDHPHTTHVISDIQ 136
Db 107 L-----VPALQAFARLHPALQVQLSTSSLIHADLIGERFDVAIRGLRLEDSTHHAVQLA 159
QY 137 DFVV 140
Db 160 SFEV 163

RESULT 49
P71763 PRELIMINARY; PRT; 339 AA.
AC P71763;
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein RV1482c.
GN RV1482C OR MT1529 OR MTCY277.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COC 1551 / Oshkosh;
RA Fleisichmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO THE GRPE FAMILY.
DR EMBL: Z79701; CAB02032.1; -.
DR EMBL: AE007022; AA845794.1; ALT_INIT.
DR PIR: E70710; E70710.
DR TIGR: MT1529; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 339 AA; 38290 MW; A0526C082B219D18 CRC64;

Query Match      8.1%; Score 86; DB 16; Length 339;
Best Local Similarity 28.3%; Pred. No. 59;
Matches 34; Conservative 14; Mismatches 50; Indels 22; Gaps 3;

QY 33 IDDAIAAEQSEITDPMKVPDHPADKFERHVGIVDFKGEIARNIEARGLKQMKR-----86

KW DNA-binding; Transcription regulation; Complete proteome.
.SQ SEQUENCE 303 AA; 33062 MW; F8C201582A0BF94 CRC64;

Query Match      8.1%; Score 86; DB 5; Length 1661;
Best Local Similarity 20.7%; Pred. No. 4.8e+02;
Matches 46; Conservative 41; Mismatches 67; Indels 68; Gaps 10;

QY 15 VSADPIHVDKITEINKAIDDAIAAEIQSEITDPMKVP-----DHADKFERHVGIVDFK 69
Db 604 VNGPEIIIEVKEBEIKKQVEDGIKE-NDTEGNDKVGKEIIESEVKEEIKKHV----- 655
QY 70 ELAMENIEARGLKQMKRQGDANVKEEGI-----VKAHLIGVHD-----D 110
Db 656 -----EEGIKENNTGNDKVGKEIIESEVKEEIKKHVEEIKKHEGKENDTESKDVIGQE 707
QY 111 IVSMYDIAVKGLDHPHTTHVISDIQDFVVALSLEISDEGNITMTSFEVQFANVNHIG 170
Db 708 IITEVKEGKIKEND-----TENKDKVIGPEM-ITEE-----VKKEIEKQEKGNKEN 753
QY 171 GLSILDPFGVLSVLTAFIQTIVRKENTKVLAPAFKRELEK 212
Db 754 ILEIKDIVIG-----QEVIIIEVKVI-----KKKVEK 781

RESULT 51
Q8KV93
```

```
ID AC Q8KV93 PRELIMINARY; PRT; 186 AA.
AC Q8KV93;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF437131; AAM49392.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR02423; Cpn60/ICP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR ATP-binding; Chaperone.
FT NON_TER 1 186
FT SEQUENCE 186 AA; 19733 MW; 8BA555908561D40A CRC64;
SQ
Query Match
Best Local Similarity 8.0%; Score 85.5; DB 2; Length 186;
Matches 41; Conservative 26; Mismatches 64; Indels 39; Gaps 8;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDPKGEIA 72
Db 16 VTAGASPI-----DLKRGIDKAVAAVVAELKAHSEVGGDDYSKVEQ-VGTVSANDSA 67

QY 73 MRNIEARGLKQMKRQGDANY---KGEIGIVKAHLIGVHDDIVSMYDLYAYKLGDLHPTT 129
Db 68 IGKLIADAMSKVKKGDTVTVEAKGTETEVK-----VEGQPFDRGY---ISPY 114

QY 130 HVSIDQDFVVALSLEIS-DEGNITWTSFEVRQFANVNVHIGLSILDPI 178
Db 115 MTNSD-----KMEASLDSPYVLITDKKISSMKDL-----LPILPEI 150

RESULT 52
Q8KVA8 PRELIMINARY; PRT; 186 AA.
AC Q8KVA8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
throughput sequencing."
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF437131; AAM49374.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR02423; Cpn60/ICP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCF1; 1.
DR ATP-binding; Chaperone.
KW
```

```
FT NON_TER 1 1
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 19772 MW; 48F171D1BB55CB17 CRC64;

Query Match
Best Local Similarity 8.0%; Score 85.5; DB 2; Length 186;
Matches 38; Conservative 30; Mismatches 64; Indels 37; Gaps 7;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDPKGEIA 72
Db 16 VTAGANPW-----DLKRGIDKAVAAVVAELKAHSEVGGDDYSKVEQ-VGTVSANDAT 67

QY 73 MRNIEARGLKQMKRQGDANY---KGEIGIVKAHLIGVHDDIVSMYDLYAYKLGDLHPTT 129
Db 68 IGKLIADAMSKVKKGDTVTVEAKGTETEVK-----VEGQPFDRGY---ISPY 114

QY 130 HVSIDQDFVVALSLEIS-DEGNITWTSFEVRQFANVNVHIGLSILDPI 178
Db 115 MTDSKMEAVL-----DNASVLITDKKISSMKDL-----MPILPEI 150

RESULT 53
Q86YC9 PRELIMINARY; PRT; 263 AA.
AC Q86YC9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042924; AAH42924.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 263 AA; 29327 MW; E9A8FAA1C6A9CB00 CRC64;

Query Match
Best Local Similarity 8.0%; Score 85.5; DB 4; Length 263;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHYDKI--TEINKAIDDAIAAEQSETIDPMK-----VPDADKFERHV-----GIVDF 67
Db 31 LHRDTTQTTLQPSVKDGLIYEDSPLVKAIVKLGHLVWDEADKAPTNTVCILKTIVE- 89

QY 68 KGEIAM---RNIEARGLKQMKRQGDANYKGEIGIVKAH-----LLIGVHDDIVSMYDLA 119
Db 90 NGEMLADGRIVA-----NSANVNGENVVVIHPDFRMIVLANRPGFPLGNDFF 140

QY 120 YKLGDLHPTTHISD-----IQDFVALS--LEISDEGNITWTS 156
Db 141 GTLGDI-F-SCHAVDNPXPHSEMLRQYGNVPPEILQKLVAAFGELSLADQGIINY-P 198

QY 157 FEVRQFANVNVHIGLSILDPIFGVSLVLTAFQ-DTVRKEMTKVL 202
Db 199 YSTREVNIVXHLQKF-----PTEG-LSSVVRNVDFDSYNNDMREIL 240

RESULT 54
Q7Y444 PRELIMINARY; PRT; 264 AA.
AC Q7Y444;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gp22 prohead core scaffold protein.
OS Enterobacteria phage RB49.
```



```

01-OCT-2002 (TReMBLrel. 22, Created)
01-OCT-2002 (TReMBLrel. 22, Last sequence update)
01-OCT-2003 (TReMBLrel. 25, Last annotation update)
Hypothetical protein (Fragment).
DKFZP761B0324
Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Amigdala;
RA Blocker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834181; CAD38878.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR SMART; SM00382; AAA; 2.
KW Hypothetical protein; ATP-binding.
FT NON_TER
SQ
SEQUENCE 745 AA; 83522 MW; F8B818BD30CC238 CRC64;

Query Match 8.0%; Score 85.5; DB 4; Length 745;
Best Local Similarity 22.0%; Pred. No. 1.8e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHVDKI--TEINKAIDDAIAIEQSETIDPMK-----VPDHADKFERHV-----GIYDF 67
Db 513 LHRDTVTQTLTLPQSVKDGILVYEDSPLVKAVKGLHILVDEADKAPTNTCILKTLVE- 571
QY 68 KGEIAM--RNTAEARGLKQKQGDANVKGEIGVKAH-----LLIGVHDDIVSMEYDLA 119
Db 572 NEMILADGRIVA-----NSANVNGRENVVVIHDFRMIIVLANRPGFFIGNDF 622
QY 120 YKLGDLHPHTHVISD-----IQFVVVALS--LEISDEGNITWTS 156
Db 623 GTLGDIF-SCHAVDNPKHSELMRQYGNVPPEPIQLKLVAFGLSLADQGIINTY-P 680
QY 157 FEVRQFANVNVNHIGLSILDPIFGVLSVLTALFQ-DTVRKEMTKVL 202
Db 681 YSTREVVNIYKHLQKF-----PTEG-LSSVVRNVFDFSYNDMDREIL 722

RESULT 57
Q7UYX3 PRELIMINARY; PRT; 807 AA.
AC Q7UYX3;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Copper-transporting ATPase (EC 3.6.3.4).
GN RB334
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12893416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleutner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294133; CAD71518.1; -.
KW Hydrolyase; Complete proteome.
SQ
SEQUENCE 807 AA; 85737 MW; 40D958BEE29484C4 CRC64;

Query Match 8.0%; Score 85.5; DB 16; Length 807;
Best Local Similarity 24.3%; Pred. No. 2e+02;
Matches 44; Conservative 32; Mismatches 69; Indels 23; Gaps 8;

QY 27 BEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVGIYDFKGLAMR-NIEARGLQMK 85
Db 429 ENLQAIQDD-----KQIQTL-----PNVALDWARLCYVLDMTNEIDERTHEA-AMAKIH 477
QY 86 RQGDANVKGEIGVKAH-----LLIGVHDDIVSMEYDLAYKLGD--LHPTTHVISDIQDFV 139
Db 478 RHFKATVGEGEGEKAHWTAAQLNWNQDDHAINLLAEQQIDDDKLPW---LLSRWRETY 534
QY 140 VALSLEISDEGNITWTSFEVRQFANVNVNHIGLSILDPIFGVLSDLVT 187
Db 535 TKRS--IGFRGRETLDKMLPRLDELINQSPSPAFAFPVSKVLEQILT 580
```

```

Matches 54; Conservative 30; Mismatches 97; Indels 41; Gaps 9;

QY 8 AAVAFVAVSADPIHYDKITEEINKAIDDAIAIEQSETIDPMKVPD---HADKFERHVI 64
Db 601 ATVVYVAVDRSLAAITAIKDIKSPKSTPDALQALSLGLRVMLTGDARSTAEAVAKQLGI 660
QY 65 VDFKGLAVENIARG--LKQMKRQGDANVKGEIGVKAHLLIGVHDDI-----VSME 115
Db 661 DDFRANVSP---BAKHDFVQQLKQEGKTVMCGDINDAPALAAASNVGIAMGTGTSVAME 717
QY 116 YDLAYKLGDLHPHTHVISDIDFVVALSLEISDEGNITWTSFEVRQFANVNVNHIG---GL 172
Db 718 -----SAG-----VTLVGGDLRGVVAAKQLSQKMTMRNQRLF---FAFAYNALGIPIAA 764
QY 173 SILDPIFGVLSDLVTA-----IFQTVRKEMTKVLAP 204
Db 765 GLIYPPFGILLSPMIAAAMSFSSVSIANALRLRTIQLSAP 806

RESULT 58
Q8EAY1 PRELIMINARY; PRT; 990 AA.
AC Q8EAY1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Glutamate-ammonia-ligase adenylyltransferase.
GN GLNE OR S03760.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015809; AAN56741.1; -.
DR TIGR; S03760; -.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR005190; GlnE.
DR Pfam; PF03710; GlnE; 2.
KW Ligase; Nucleotidyltransferase; Transferase; Complete proteome.
SQ
SEQUENCE 990 AA; 113313 MW; 77C64A06423F790E CRC64;

Query Match 8.0%; Score 85.5; DB 16; Length 990;
Best Local Similarity 26.2%; Pred. No. 2.7e+02;
Matches 44; Conservative 32; Mismatches 69; Indels 23; Gaps 8;

QY 27 BEINKAIDDAIAIEQSETIDPMKVPDHADKFERHVGIYDFKGLAMR-NIEARGLQMK 85
Db 429 ENLQAIQDD-----KQIQTL-----PNVALDWARLCYVLDMTNEIDERTHEA-AMAKIH 477
QY 86 RQGDANVKGEIGVKAH-----LLIGVHDDIVSMEYDLAYKLGD--LHPTTHVISDIQDFV 139
Db 478 RHFKATVGEGEGEKAHWTAAQLNWNQDDHAINLLAEQQIDDDKLPW---LLSRWRETY 534
QY 140 VALSLEISDEGNITWTSFEVRQFANVNVNHIGLSILDPIFGVLSDLVT 187
Db 535 TKRS--IGFRGRETLDKMLPRLDELINQSPSPAFAFPVSKVLEQILT 580
```

```

RESULT 59
Q72619
ID Q72619 PRELIMINARY; PRT; 1062 AA.
AC Q72619;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Eve;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Eve;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053674; AAH53674.1; -
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 1062 AA; 119150 MW; E629CB6B8B018E21 CRC64;

Query Match 8.0%; Score 85.5; DB 4; Length 1062;
Best Local Similarity 22.0%; Pred. No. 2.9e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

Qy 20 IHVDKI--TEINKAIDDAIAAEQSETIDPMK-----VPDHADKFERHV-----GIVDF 67
Db 830 LHRDTTQTTLQPSVKGLIVVEDSPLVKAVKGLHVLVVDKAPTNTVCILKTLVE- 888

Qy 68 KGEIAM--RNIEARGLKQKRGDANVKGEGIVKVAH-----LLIGVDDIVSMYDIA 119
Db 889 NGENIMADGRIVA-----NSANVNGRNVVWHPDFRMVILANRPGPFGLGNDFF 939

Qy 120 YKLGDLHPTHTVDS-----IQDFVALS--LEISDEGNTMTS 156
Db 940 GTLGDI-F-SCHAVDNPXPHSELEMLRQYGNVPEPIQLKLVAFGEIRSLADQGLINY 997

Qy 157 FEVRQFANVNHIGLSILDPFGVLSVDULTAIFQ-DTVKEMTKVL 202
Db 998 YSTREVVNIHKLQKF----PTEG-LSSVVRNVDFDYSYNNDMREIL 1039

RESULT 60
Q91250
ID Q91250 PRELIMINARY; PRT; 1154 AA.
AC Q91250;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998)
DR EMBL; AB011136; BAA25490.1; -
DR FIC; T00335; T00335.
DR InterPro; IPR002035; VWF_A.
DR SMART; SMO0327; VWA; 1.

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
GN MASS1.3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21432018; PubMed=11545713;
RA Skradski S.L., Clark A.M., Jiang H., White H.S., Pu Y., Ptacek L.J.;
RT "A novel gene causing a Mendelian audiogenic mouse epilepsy."
RL Neuron 31:537-544(2001).
DR EMBL; AF405694; AAL06014.1; -
DR MGD; MGI:1274784; Mass1.
GO; GO:0005507; F: copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 6.
DR SMART; SMO0237; Calx beta; 7.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
SQ SEQUENCE 1154 AA; 123814 MW; CBDACF4ECE3B5D06 CRC64;

Query Match 8.0%; Score 85.5; DB 11; Length 1154;
Best Local Similarity 21.4%; Pred. No. 3.3e+02;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

Qy 26 TEEINKAIDDAIAAEQSETIDPMKVPDHADKFERHV-----GIYDFKGLAMRNIEA 78
Db 882 TVTVNLANDNVAGIVSFQTSRSRVIGHEGEMLQFHVVRTPPGRGVTVNWKVVGQNVLE 941

Qy 79 RGLKMKRQGDANVKG-----EGIVKAHLIGVHDDIVSME-----YDLAYKLGDLH 126
Db 942 -----NPNFTGQLFFSEGLTKNTIFVHLDDNDNIPKEVYQVLYXDV--KTGVS 990

Qy 127 PTTHVISDIQDFVVALSLEISDEG-----NITMTS--FEVRQFANV-----NHIGLSIL 175
Db 991 PAGVALDQAQYAAVLTVESASDEPHGVNLFALSSRFVQLQEVNTIQLFVNRREGSLGAI 1050

Qy 176 DPIFGVLSVDULT 187
Db 1051 NVTYATVPGIVS 1062

RESULT 61
Q60310
ID Q60310 PRELIMINARY; PRT; 1441 AA.
AC Q60310;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein KIAA0564 (Fragment).
GN KIAA0564
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998)
DR EMBL; AB011136; BAA25490.1; -
DR FIC; T00335; T00335.
DR InterPro; IPR002035; VWF_A.
DR SMART; SMO0327; VWA; 1.

```



```
DR PROSITE; P850234; WFA; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1441 AA; 162836 MW; 360A40ECA4E3054D CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 4.4e+02;
Matches 50; Conservative 35; Mismatches 81; Indels 61; Gaps 14;

QY 20 IHVDKI--TEEINKAIDDAIAEQSETIDPMK-----VPDHADKFERHV-----GIVDF 67
Db 343 LHRDITVQTLTQPSVKDGLVVEDSPLKAVKGLHVLVDEADKAPTNTVILKTLVE- 401
QY 68 KGBLAM---RNIEARGLKMKRGDANVKGEGIVKAH-----LIGVHDDIVSMEYDLA 119
Db 402 NGEMILADGERIVA-----NSANVGRENVVVHPDFRMVLANRPGPFPLGNDF 452
QY 120 YKLGDLHPHTHVISD-----IDPFVVALS---LEISDEGNITWTS 156
Db 453 GTLGDI-F-SCHAYDNPKPHEMLRQYGNFVPEPILQKLVAAFGSLRLADQGIINY-P 510
QY 157 FEVRQFANVNVHIGGLSILDPFVGLSDVLTAFQ-DTVRKEMTKVL 202
Db 511 YSTREVVNVKHLQKF-----PTSG-LSSVRNVFDFDSYNDNRIL 552

RESULT 62
ID Q91ZS1 PRELIMINARY; PRT; 2238 AA.
AC Q91ZS1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MASS1.2.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;
RA "A novel gene causing a mendelian audiogenic mouse epilepsy.";
RL Neuron 31:537-544(2001).
DR EMBL; AF405693; AAL06013.1; -.
DR MGD; MGI:1274784; Mass1.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 12.
DR SMART; SM00237; Calx beta; 11.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00879; ODR DC 2.2; 1.
SQ SEQUENCE 2238 AA; 244054 MW; AFE2B4F8A1BDAF58 CRC64;

Query Match
Best Local Similarity 8.0%; Score 85.5; DB 11; Length 2238;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAEQSETIDPMKVPDHADKFERHV-----GIVDFKGLAMRNIEA 78
Db 1966 TVTVNLANDNVAGIVSFQTASRVIGHEGMLQFHVVRTPPGRGNTVNVKVGQNLV 2025
QY 79 RGLKMKRQGDANVKG-----EGIVKAHLIIGVHDDIVSME-----YDLAYKLGDLH 126
Db 2026 -----NFANFTGQLFFSEGLTKNTIFVHLLDDNIPKEKVVQVLYDV--KTQGS 2074
QY 127 PTHVISDIQDFVVALSLEISDEG---NITWTS-FEVRQFANVV-----NHIGGLSIL 175
```

```
Db 2075 PAGVALLDAQYAVALTVEASDEPHGVNLFALSSRFVVLQEANVTIQLFVNREFGSLGAI 2134
QY 176 DPIFGVLSDLVT 187
Db 2135 NVTYATVPGIVS 2146

RESULT 63
Q91ZS2 PRELIMINARY; PRT; 2780 AA.
ID Q91ZS2;
AC Q91ZS2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MASS1.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX Skradski S.L., Clark A.M., Jiang H., White H.S., Fu Y., Ptacek L.J.;
RA "A novel gene causing a mendelian audiogenic mouse epilepsy.";
RL Neuron 31:537-544(2001).
DR EMBL; AF405692; AAL06012.1; -.
DR MGD; MGI:1274784; Mass1.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA_like_1ec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF03160; Calx-beta; 12.
DR SMART; SM00237; Calx beta; 11.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 1.
DR PROSITE; PS00879; ODR DC 2.2; 1.
SQ SEQUENCE 2780 AA; 302875 MW; FF838B5864249ADE CRC64;

Query Match
Best Local Similarity 8.0%; Score 85.5; DB 11; Length 2780;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAEQSETIDPMKVPDHADKFERHV-----GIVDFKGLAMRNIEA 78
Db 2508 TVTVNLANDNVAGIVSFQTASRVIGHEGMLQFHVVRTPPGRGNTVNVKVGQNLV 2567
QY 79 RGLKMKRQGDANVKG-----EGIVKAHLIIGVHDDIVSME-----YDLAYKLGDLH 126
Db 2568 -----NFANFTGQLFFSEGLTKNTIFVHLLDDNIPKEKVVQVLYDV--KTQGS 2616
QY 127 PTHVISDIQDFVVALSLEISDEG---NITWTS-FEVRQFANVV-----NHIGGLSIL 175
Db 2617 PAGVALLDAQYAVALTVEASDEPHGVNLFALSSRFVVLQEANVTIQLFVNREFGSLGAI 2676
QY 176 DPIFGVLSDLVT 187
Db 2677 NVTYATVPGIVS 2688

RESULT 64
Q9VHN7 PRELIMINARY; PRT; 6298 AA.
ID Q9VHN7;
AC Q9VHN7;
DT 01-NAR-2002 (TrEMBLrel. 20, Created)
DT 01-NAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Very large G protein-coupled receptor 1.
GN MASS1 OR VLGR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21634833; PubMed=11606593;
RA McMillan D.R., Kayes-Wandover K.M., Richardson J.A., White P.C.;
RT "Very Large G Protein-coupled Receptor-1, the Largest Known Cell
RT Surface Protein, is Highly Expressed in the Developing Central Nervous
RT System.";
RL J. Biol. Chem. 277:785-792(2002).
DR EMBL; AF435926; AAL30812.1; -.
DR MGD; MGI:1274784; Maesi.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0004930; F:G-protein coupled receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0007219; P:neuropeptide signaling pathway; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR009039; EAR.
DR InterPro; IPR005492; EPTP.
DR InterPro; IPR000832; GPCR secretin.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm2; 1.
DR Pfam; PF03160; Calx-beta; 19.
DR Pfam; PF03736; EPTP; 1.
DR SMART; SM00237; Calx beta; 20.
DR SMART; SM00303; GPS; 1.
DR PROSITE; PS00912; EAR; 4.
DR PROSITE; PS02221; GPS; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
KW Receptor.
SQ SEQUENCE 6298 AA; 687437 MW; F99E731B0ADB66D CRC64;

Query Match      8.0%; Score 85.5; DB 11; Length 6298;
Best Local Similarity 21.4%; Pred. No. 3.1e+03;
Matches 41; Conservative 32; Mismatches 78; Indels 41; Gaps 8;

QY 26 TEEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHV-----GIVDFKGLAMRNIEA 78
Db 2686 TVTVNILLANDVAGIVSFQATSRVIGHEGEMLQFHVVRTPPGRGNTVANKVVGQNLV 2745
QY 79 RGLQMKRQGDANVKG-----EGIVKAHLIGVHDDIVSVE-----YDLAYKGLDLH 126
Db 2746 -----NFANFTGQLFFSEGLTKNTIFVHLDDNIPKEEYQVYVLYDV--KTQGV 2794
QY 127 PTHVVISDIPQVVALSLEISDEG---NITMTS-FEVROFANVV-----NHIGGLSIL 175
Db 2795 PAGVALLDAQGYAAVLIVASDEPHGVNLFALSSRFVVLQEVANTIQLFVNRFGSLGAI 2854
QY 176 DPIFGVLSDLVT 187
Db 2855 NVTYATVPGIVS 2866

RESULT 65
Q8KVK6 PRELIMINARY; PRT; 185 AA.
AC Q8KVK6;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
RT throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437097; AAM49358.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR ATP-binding; Chaperone.
DR NON_TER 1
DR NON_TER 186
SQ SEQUENCE 186 AA; 20188 MW; D798608CB07C9E7C CRC64;

Query Match      8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVIGVDFKGLA 72
Db 16 VAAGANPM-----DLKRGIDKAVAKVVEGIKAQAEVGEFDEKIE-NVARISANDSV 67
QY 73 MNIEARGLKQMKRQGDANVKGEGIVKAHLIGVHDDIV-SMEYDLAY 120
Db 68 IGELIAEAMKKVRDGVITVEDAKG-----TDTYVDVVEGQMQRG 109

RESULT 67
Q8KVQ2 PRELIMINARY; PRT; 186 AA.
ID Q8KVQ2
```

```
RT "Extensive profiling of a complex microbial community by high-
RT throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF436976; AAM49237.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR ATP-binding; Chaperone.
DR NON_TER 1
DR NON_TER 185
SQ SEQUENCE 185 AA; 20485 MW; E9A9FC0397C5D322 CRC64;

Query Match      8.0%; Score 85; DB 2; Length 185;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVIGVDFKGLA 72
Db 16 VAAGANPM-----DLKRGIDKAVAKVVEGIKAQAEVGEFDEKIE-NVARISANDSV 67
QY 73 MNIEARGLKQMKRQGDANVKGEGIVKAHLIGVHDDIV-SMEYDLAY 120
Db 68 IGELIAEAMKKVRDGVITVEDAKG-----TDTYVDVVEGQMQRG 109

RESULT 66
Q8KVC0 PRELIMINARY; PRT; 186 AA.
ID Q8KVC0;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-
RT throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437097; AAM49358.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR ATP-binding; Chaperone.
DR NON_TER 1
DR NON_TER 186
SQ SEQUENCE 186 AA; 20188 MW; D798608CB07C9E7C CRC64;

Query Match      8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;

QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVIGVDFKGLA 72
Db 16 VAAGANPM-----DLKRGIDKAVAKVVEGIKAQAEVGEFDEKIE-NVARISANDSV 67
QY 73 MNIEARGLKQMKRQGDANVKGEGIVKAHLIGVHDDIV-SMEYDLAY 120
Db 68 IGELIAEAMKKVRDGVITVEDAKG-----TDTYVDVVEGQMQRG 109

RESULT 67
Q8KVQ2 PRELIMINARY; PRT; 186 AA.
ID Q8KVQ2
```

AC Q8KV2; 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
Crosby W.L., Hemmingsen S.M.;
RT Extensive profiling of a complex microbial community by high-throughput sequencing.;
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF437042; AAM49303.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1 1
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 20151 MW; 7BE9B5B3329114D CRC64;

Query Match 8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;
QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSEITIDPMKVPDHDKFERHVGIVDFKGLA 72
DB 16 VAAGANPM-----DLKRGIDKAVKVEGIGKAAQAEVGEDEFDKIE-NVARISANDSV 67
QY 73 MRNIEARGLKQKRGDANVKGEIGVKAHLIGVHDDIV-SMEYDLAY 120
DB 68 IGEIAEAMKVKRDGVITVEDAKG-----TDTYVDVVEGMOFDRGY 109

RESULT 68
Q8KHQ8 PRELIMINARY; PRT; 186 AA.
AC Q8KHQ8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DE 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 60 kDa chaperonin (Fragment).
GN CPN60.
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
Crosby W.L., Hemmingsen S.M.;
RT Extensive profiling of a complex microbial community by high-throughput sequencing.;
RL Appl. Environ. Microbiol. 68:3055-3066 (2002).
DR EMBL; AF436897; AAM49158.1; -
DR EMBL; AF437007; AAM49268.1; -
DR EMBL; AF437129; AAM49390.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1 1
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 20201 MW; 7BE9B5B526290CEF CRC64;

Query Match 8.0%; Score 85; DB 2; Length 186;
Best Local Similarity 23.9%; Pred. No. 32;
Matches 26; Conservative 23; Mismatches 44; Indels 16; Gaps 4;
QY 13 VAVSADPIHYDKITEINKAIDDAIAAEQSEITIDPMKVPDHDKFERHVGIVDFKGLA 72
DB 16 VAAGANPM-----DLKRGIDKAVKVEGIGKAAQAEVGEDEFDKIE-NVARISANDSV 67
QY 73 MRNIEARGLKQKRGDANVKGEIGVKAHLIGVHDDIV-SMEYDLAY 120
DB 68 IGEIAEAMKVKRDGVITVEDAKG-----TDTYVDVVEGMOFDRGY 109

RESULT 69
Q8UZZ6 PRELIMINARY; PRT; 292 AA.
AC Q8UZZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DE 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PYRAB10690.
GN PYRAB10690 OR PAB0711.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248286; CAB49980.1; -
DR PIR; G75084; G75084.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 292 AA; 31578 MW; AE410B3578BFA8C CRC64;

Query Match 8.0%; Score 85; DB 17; Length 292;
Best Local Similarity 20.7%; Pred. No. 59;
Matches 44; Conservative 31; Mismatches 62; Indels 76; Gaps 7;
QY 29 INKAIDD-----AIAAEQSEITIDPMKVPDHDKFERHVGIVDFK----- 68
DB 11 INKVIENPKVIIFPLIVLIPSVPLAFLEKESNIKPLNFEEAGVIEKHGAISDMKLPNL 70
QY 69 -----GELAMRNTAE-----RGLKQMKRGDANVKGEIGVKAHLIGVHDDIVSMY 116
DB 71 KLLLVGLQLLSAVQYSIIHYVKTGSTNGEAFKGLGVNIQMLL----- 118
QY 117 DLAYKLGDLHPTTHVISDIDQFVVVALSLEISDEGNITWTSPFVRQF-ANVVNHIGGL--S 173
DB 119 -----NVISTL-----IVLVAFLIAIFPPMIIVGIGILES 149
QY 174 ILDPFGVLSDLTAIFQDITVRKEMTKVLAPAF 206
DB 150 TGAVIFGILLVILSLIGLSAIGTMTSIVPEAY 182

RESULT 70
Q8RDZ2 PRELIMINARY; PRT; 401 AA.
AC Q8RDZ2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DE 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ABC transporter permease protein.
GN FNI349.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;

DE Hemoisocitrate dehydrogenase.
GN HICD.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermoales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyazaki J., Kobashi N., Nishiyama M., Yamane H.;
RT "Characterization of homoisocitrate dehydrogenase from Thermus
thermophilus HB27.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AS075751; BAB8861.1; -
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001804; Isohdh.
DR Pfam; PF00180; Isohdh; 1.
SQ SEQUENCE 334 AA; 35922 MW; 82E018FD744FB49 CRC64;

Query Match 7.9%; Score 84.5; DB 2; Length 334;
Best Local Similarity 22.8%; Pred. No. 77;
Matches 50; Conservative 34; Mismatches 94; Indels 41; Gaps 7;
QY 10 VAFVAVSDPIHYDKITEINKAIDAIAAEIQESTIDPMKVPDHAKFERHGVIVFKG 69
Db 114 LVIVRENTGLVEQERRYLDVAIADAVISKASERI-----G 151
QY 70 ELAMENIEARGLKQMKGQDNV-----KGEIGVKAHLIGVHDDIV---SMEYD 117
Db 152 RAALRIARERPKTHIAHKANVLPTQLGLFDTVKVAKDFPLVNVDIIIVDCANQLV 211
QY 118 LAYKLGDLLHTTHVISDI-QDFVVAL--SLSEISDEGNITMTSFVRQPANVVNHGGLSI 174
Db 212 MRPERFDVIIVTNLLGDLILSLAAGLVGGLAPSNGIGDTTAVFEPVHGSAPIAGKI 271
QY 175 LDPIFGVLSDLTAIFQDTV-RKENTKVLAFAFRKLEK 212
Db 272 ANPATAILS---AAAMDLYLGEKAARKVEXAVDLVLER 307

RESULT 75
Q8AOK9 PRELIMINARY; PRT; 426 AA.

ID Q8AOK9
AC Q8AOK9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC oligo/dipeptide transport, ATP-binding protein.
GN B4012.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12863928;
RA Xu J., Bjursell M.K., Hamrod J., Deng S., Carmichael L.K.,
Chiang H.C., Hooper I.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
Science 299:2074-2076(2003).
RL EMBL; AF016943; AAO79117.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 426 AA; 48748 MW; 635D2514FF7FEF8 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 426;
Best Local Similarity 23.4%; Pred. No. 1.1e+02;
Matches 44; Conservative 36; Mismatches 67; Indels 41; Gaps 9;
DE Hemoisocitrate dehydrogenase.
GN HICD.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermoales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyazaki J., Kobashi N., Nishiyama M., Yamane H.;
RT "Characterization of homoisocitrate dehydrogenase from Thermus
thermophilus HB27.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AS075751; BAB8861.1; -
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001804; Isohdh.
DR Pfam; PF00180; Isohdh; 1.
SQ SEQUENCE 334 AA; 35922 MW; 82E018FD744FB49 CRC64;

Query Match 7.9%; Score 84.5; DB 2; Length 334;
Best Local Similarity 22.8%; Pred. No. 77;
Matches 50; Conservative 34; Mismatches 94; Indels 41; Gaps 7;
QY 10 VAFVAVSDPIHYDKITEINKAIDAIAAEIQESTIDPMKVPDHAKFERHGVIVFKG 69
Db 114 LVIVRENTGLVEQERRYLDVAIADAVISKASERI-----G 151
QY 70 ELAMENIEARGLKQMKGQDNV-----KGEIGVKAHLIGVHDDIV---SMEYD 117
Db 152 RAALRIARERPKTHIAHKANVLPTQLGLFDTVKVAKDFPLVNVDIIIVDCANQLV 211
QY 118 LAYKLGDLLHTTHVISDI-QDFVVAL--SLSEISDEGNITMTSFVRQPANVVNHGGLSI 174
Db 212 MRPERFDVIIVTNLLGDLILSLAAGLVGGLAPSNGIGDTTAVFEPVHGSAPIAGKI 271
QY 175 LDPIFGVLSDLTAIFQDTV-RKENTKVLAFAFRKLEK 212
Db 272 ANPATAILS---AAAMDLYLGEKAARKVEXAVDLVLER 307

RESULT 75
Q8AOK9 PRELIMINARY; PRT; 426 AA.

ID Q8AOK9
AC Q8AOK9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC oligo/dipeptide transport, ATP-binding protein.
GN B4012.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12863928;
RA Xu J., Bjursell M.K., Hamrod J., Deng S., Carmichael L.K.,
Chiang H.C., Hooper I.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
Science 299:2074-2076(2003).
RL EMBL; AF016943; AAO79117.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 426 AA; 48748 MW; 635D2514FF7FEF8 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 426;
Best Local Similarity 23.4%; Pred. No. 1.1e+02;
Matches 44; Conservative 36; Mismatches 67; Indels 41; Gaps 9;
DE Hemoisocitrate dehydrogenase.
GN HICD.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermoales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyazaki J., Kobashi N., Nishiyama M., Yamane H.;
RT "Characterization of homoisocitrate dehydrogenase from Thermus
thermophilus HB27.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AS075751; BAB8861.1; -
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001804; Isohdh.
DR Pfam; PF00180; Isohdh; 1.
SQ SEQUENCE 334 AA; 35922 MW; 82E018FD744FB49 CRC64;

Query Match 7.9%; Score 84.5; DB 2; Length 334;
Best Local Similarity 22.8%; Pred. No. 77;
Matches 50; Conservative 34; Mismatches 94; Indels 41; Gaps 7;
QY 10 VAFVAVSDPIHYDKITEINKAIDAIAAEIQESTIDPMKVPDHAKFERHGVIVFKG 69
Db 114 LVIVRENTGLVEQERRYLDVAIADAVISKASERI-----G 151
QY 70 ELAMENIEARGLKQMKGQDNV-----KGEIGVKAHLIGVHDDIV---SMEYD 117
Db 152 RAALRIARERPKTHIAHKANVLPTQLGLFDTVKVAKDFPLVNVDIIIVDCANQLV 211
QY 118 LAYKLGDLLHTTHVISDI-QDFVVAL--SLSEISDEGNITMTSFVRQPANVVNHGGLSI 174
Db 212 MRPERFDVIIVTNLLGDLILSLAAGLVGGLAPSNGIGDTTAVFEPVHGSAPIAGKI 271
QY 175 LDPIFGVLSDLTAIFQDTV-RKENTKVLAFAFRKLEK 212
Db 272 ANPATAILS---AAAMDLYLGEKAARKVEXAVDLVLER 307

RESULT 75
Q8AOK9 PRELIMINARY; PRT; 426 AA.

ID Q8AOK9
AC Q8AOK9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC oligo/dipeptide transport, ATP-binding protein.
GN B4012.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12863928;
RA Xu J., Bjursell M.K., Hamrod J., Deng S., Carmichael L.K.,
Chiang H.C., Hooper I.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
Science 299:2074-2076(2003).
RL EMBL; AF016943; AAO79117.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 426 AA; 48748 MW; 635D2514FF7FEF8 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 426;
Best Local Similarity 23.4%; Pred. No. 1.1e+02;
Matches 44; Conservative 36; Mismatches 67; Indels 41; Gaps 9;
DE Hemoisocitrate dehydrogenase.
GN HICD.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermoales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyazaki J., Kobashi

```

QY 31 KAIDDAIAAI-EQGETIDPMKVPDPHADKFERHVGI-----VDFKGLAMRNIEARGLKQMKR 86
DB 184 KAMDDLVOEVTKKETANQAKIL-YLKVDKIAIKGVNFKGIQLDLNVVAEGNKKKN 242
QY 87 QGDANVKEGEGIV-----KAHLIGVHDDIVSMYDLYAYKLDLHPHTHVISDIQDF 138
DB 243 RITAYTEDKENILFKVAILKSKAFVL-----DFVDVTLPCS-TLMEL-----VTKRVPAP 291
QY 139 VVALSLEISDSGNITMTSFEVROFANVNV-----HIGLSILDPIFGVL 182
DB 292 IYPSIVILD-GDVMMKNDRKINNADNILLPKNKSPERLLASYLNSDVLWSKI 350
QY 183 SDVLTAF 190
DB 351 ADGYTKQF 358

RESULT 76
Q8PG51 PRELIMINARY; PRT; 710 AA.
AC Q8PG51;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Chemotaxis protein.
GN XAC3768.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AF012026; AAM38611.1;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chttaxis_transd.
DR InterPro; IPR003660; HAMT.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMT; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTXNSDCR.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDCUC_2; 1.
DR PROSITE; PS50885; HAMT; 2.
KW Complete proteome.
SQ SEQUENCE 710 AA; 75422 MW; D84CT7536SDC03PD0 CRC64;

```

```

Query Match 7.9%; Score 84.5; DB 16; Length 710;
Best Local Similarity 23.1%; Pred. No. 2.16e+02;
Matches 54; Conservative 31; Mismatches 80; Indels 69; Gaps 11;

```

```

QY 25 ITTEINKAI-----DDAIAAIQSE-----TIDPMK-----VFDHADKPERH 61
DB 217 ITQPLNRAVGIAHAYSSGKLDNAISIDRENDEVQLQAMDSMORQLRRVIAQOTEMGEKH 276
QY 62 -VGIV-----DFKGLAMRNIEARGLKQMKRQGDANVKEEGIVKAHLIGVHDDIVS 113
DB 277 EAGVVSRIAEQEFPEFG-----SMVRASNALVAGHIAVTWQIVALT 319
QY 114 MEY---DLAYKLGDLHPHTHVISDIQDFVVALSLEISDE-----GNITM-----TS 156
DB 320 ERYAAGDLSQQLPELPGEKAVISQSLNDVRASLLSINGEIKELASAAAAGDFTARGDETA 379
QY 157 FE--VRQFANVNVHIGLSILDPIFGVLSVLTAFQDFTVKEMTKVLAPAKR 208
DB 380 FEHDFRQIVVDLNLQ--NSTSDQSIGALSQVLQATIASGDLTARKHGNFAGVFAR 431

RESULT 77
Q8CQ88 PRELIMINARY; PRT; 817 AA.
AC Q8CQ88;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Endopeptidase.
GN SE0287.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016744; AAC03884.1;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004513; F:nuclease activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003959; AAA ATPase centr.
DR InterPro; IPR001270; Chaprtnin_clpA/B.
DR InterPro; IPR004176; Clp_N.
DR InterPro; IPR001943; UVRE/C.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp_N; 1.
DR Pfam; PF02151; UVR; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00570; CLPAB_1; 1.
DR PROSITE; PS00371; CLPAB_2; 1.
DR PROSITE; PS50151; UVR; 1.
KW Complete proteome.
SQ SEQUENCE 817 AA; 91012 MW; 76E6455923462294 CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 817;
Best Local Similarity 23.0%; Pred. No. 2.5e+02;
Matches 50; Conservative 35; Mismatches 87; Indels 45; Gaps 11;

QY 18 DPH-----YDKITEEINKAIDDAIAAEQSETIDPMKVPDPHADKFERHVGIVDFKGLA 72
DB 504 DTLHRRVIGONDVANSISKAVRRARAGLK-----DP-----KRPIGSFILGPTG 548
QY 73 MNIE-ARGL-KQMKRQGDANVKEG-EGIVKAHL---LIGV-----HDDIVSMYELAY 120
DB 549 VKTELARALAEAFEGEDAMIRVDMSEFMKHAHSRLVGAPGVGHDDGGQITKVR 608
QY 121 KGLDLHPHTHVISD-----IQDFVVALSLEISDEGNITMTSFEVROFANV-----NHIGGL 172
DB 609 K-----PYSVILFOIEKAHPDVFENILLQVLDLGHLDTKGRTVDFTNTVIIMTSNVAQ 663

```

```
QY 173 SILDPIFGVLSVLTAFQDVTVRKEMTKVLAPAKKE 209
Db 664 ELQDQRFAGGASGSDYETVRKTKMKLKNFRPE 700

RESULT 78
Q898W1 PRELIMINARY; PRT; 841 AA.
AC Q898W1
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE Protein translocase subunit secA.
GN SECA OR CTG00327.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / B88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR EMBL; AEO15937; AAC34968.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006605; P:protein targeting; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000185; SecA.
DR InterPro; IPR004027; SEC_C motif.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF02810; SEC-C; 1.
DR Pfam; PF01043; SecA protein; 1.
DR PRINTS; PR00906; SECA.
DR PROSITE; PS01312; SECA; 1.
DR Complete proteome.
KW SEQUENCE 841 AA; 96627 MW; 183803C31DB55FCA CRC64;

Query Match 7.9%; Score 84.5; DB 16; Length 841;
Best Local Similarity 22.8%; Pred. No. 2.6e+02;
Matches 60; Conservative 37; Mismatches 67; Indels 99; Gaps 16;

QY 23 DKITEINKA--TDDAIAAEQSETIDPM---KVPD---HADKFRHVGIVDFKGLAM 73
Db 425 DEIVETHKGPVLVGTVSTSEKSELISEMLKRGVPHVNLNXYHKEAEIVSHAGEYGN 484
QY 74 ----RNIEARGLKMRQGDANYKGBEGIVKAH--LLIGV--HDD-----110
Db 485 VTATNAGRG-----TDKLEEVKAGGLKIIGTERHESRRIDNQLRGRSGRQ 535
QY 111 -----IVSMEVDL-----AYKGLDHPHTHVISD---IQDFVVALSLEISD 148
Db 536 DPGASRFVSLDDLMRIFGSKLKGIVEKLG-----LGDDAEIESKQVSNATENAQ 587
QY 149 ---EGNITMTSFEVR---OFANVNH-----IGGLSILDPFGVLSDLVLTAI 189
Db 588 KQVEGN---NFDIRKTLIQYDDVINKQRIIYKQSEVLEGADLKQIQCEMIRDVINSV 643
QY 190 FQDVTVRKEMTKVLAPAKRLEK 212
Db 644 VDHSID-----IEEFPKEELDK 661

RESULT 79
Q89XK3 PRELIMINARY; PRT; 1429 AA.
AC Q89XK3
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE LD43558p.
GN RPI1215 OR CG1554.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B. Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089555; AAL90293.1; -.
DR FlyBase; FBGN0003277; RPI1215.
DR GO; GO:0005665; C:DNA-directed RNA polymerase II, core complex; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR006592; RNA_pola_N.
DR InterPro; IPR000684; RNA_poliI_repeat.
DR InterPro; IPR000722; RNA_poliA.
DR InterPro; IPR007066; RNA_poliRpb1_3.
DR InterPro; IPR007083; RNA_poliRpb1_4.
DR InterPro; IPR007081; RNA_poliRpb1_5.
DR InterPro; IPR007075; RNA_poliRpb1_6.
DR InterPro; IPR007073; RNA_poliRpb1_7.
DR Pfam; PF00623; RNA_poliRpb1_2; 1.
DR Pfam; PF04983; RNA_poliRpb1_3; 1.
DR Pfam; PF05000; RNA_poliRpb1_4; 1.
DR Pfam; PF04998; RNA_poliRpb1_5; 1.
DR Pfam; PF04992; RNA_poliRpb1_6; 1.
DR Pfam; PF04990; RNA_poliRpb1_7; 1.
DR Pfam; PF05001; RNA_poliRpb1_R; 22.
DR SMART; SM00663; RPOLA_N; 1.
DR PROSITE; PS00115; RNA_POLI; REPEAT; 11.
DR SEQUENCE 1429 AA; 157524 MW; 540A045520CAEDA6 CRC64;

Query Match 7.9%; Score 84.5; DB 5; Length 1429;
Best Local Similarity 22.7%; Pred. No. 5.3e+02;
Matches 58; Conservative 33; Mismatches 86; Indels 79; Gaps 12;

QY 17 ADPIHYKITEINKAIDDAIAAEQSETIDPMKVPDHA--DKERHVGIV-----DFKG 69
Db 228 ADPQTYNEIQQAIAKAKDDVINVIQKARNMELEPTPGTLENKVNRLNDARDKGT 287
QY 70 ELAMRNI--EARGLKMRKQ-----GDANKGEE---GIVKAHLLIGVHD 109
Db 288 GSAKSLTEYNLNKAMVSGSGKSNINISQVIACGQGVNVEGKEIPYGRKRTLPFIKD 347
QY 110 D-----IVSMEVDLAYKGLDHPHTT---HVISDIQDFVVALSLEISDEGNI-----T 153
Db 348 DYGPESRGFVNSY-----LAGLTFSEFFYHAMGG--REGLIDTAVKTAETGYIQRLIKA 401
QY 154 MTSFEVRQFANVNHIGLSIL---DPFGVLSD-----184
Db 402 MESVMVNDGVTRNSVGQIIQIRYGEDGLCGELVEFQNMPTVKLSNKSFEKPFKEDWSNE 461
QY 185 -VLTAIFQDVTVRKEMT 199
Db 462 RLMKKVFTDDVIKEMT 477
```

RESULT 80
Q8KV19 PRELIMINARY; PRT; 186 AA.
AC Q8KV19; 186 AA.
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).
DR EMBL; AF437230; AAM49491.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60 TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20429 MW; CEC0EAEC4A1A7FB CRC64;

Query Match 7.9%; Score 84; DB 2; Length 186;
Best Local Similarity 20.8%; Pred. No. 39;
Matches 40; Conservative 35; Mismatches 63; Indels 54; Gaps 8;
QY 13 VAVSADPHYDKITEINKATDDAIAAEQSETIDPMKVPDHDADKPERHVGIVDF 72
Db 16 VAGANPM-----DLKRGIDKAVAVVAFIKHAEQVD-----DNYDKIEQ-VATVSA 67
QY 73 MRNIEARGLKMKQKQGDANVKGEIGVKAHLITGVHDDIVSMYDLYAYKLGDLHPTTH-- 130
Db 68 IGRLLAEAFKVKSGDGVITIEESK--SRETYIGV--VEGQDFDRGYISAYMAT 121
QY 131 -----VISDIQDF-----VVALSLEISDGNITMTSFEVRQFA 163
Db 122 ECQVDNPHYLLYDKKISNKEFLFLOPAESGRPLLVIAEDVDSEALTTL----- 172
QY 164 NVNHHI-GGLSI 174
Db 173 -VNRLRGGLKI 183

RESULT 81
Q8KV15 PRELIMINARY; PRT; 186 AA.
AC Q8KV15; 186 AA.
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE 60 kDa chaperonin (Fragment).
GN CPN60
OS uncultured pig faeces bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=190966;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22034982; PubMed=12039767;
RA Hill J.E., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G.,
RA Crosby W.L., Hemmingsen S.M.;
RT "Extensive profiling of a complex microbial community by high-throughput sequencing.";
RL Appl. Environ. Microbiol. 68:3055-3066(2002).

DR EMBL; AF437010; AAM49271.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60 TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20276 MW; 9B939C6DBCDF8F78 CRC64;

Query Match 7.9%; Score 84; DB 2; Length 186;
Best Local Similarity 23.0%; Pred. No. 39;
Matches 43; Conservative 35; Mismatches 65; Indels 44; Gaps 10;

QY 13 VAVSADPHYDKITEINKATDDAIAAEQSETIDPMKVPDHDADKPERHVGIVDF 67
Db 16 VAGANPM-----DLKRGIDKAVAVVAFIKHAEQVD-----DNYDKIEQ-VATVSA 62
QY 68 KGLAMRNIEARGLKMKQKQGDANVKGEIGVKAHLITGVHDDIVSMYD-----LAYKLG 123
Db 63 NNDPEIGKLLADAMKVKSGDGVITIE-ESKTMQTEL-----DLVEGQDFDRGYISAYMAT 116
QY 124 DLHP-----TTHVISDIQDFVVALSLEISDGNITMTSFEVRQFA--NVNHH 168
Db 117 DMQRMEAVLDDPYLLITDKKISNKEFLFLOPAESGRPLLVIAEDVDSEALTTLVNR 176
QY 169 I-GGLSI 174
Db 177 LRGLGLKI 183

RESULT 82
Q8DWU5 PRELIMINARY; PRT; 306 AA.
AC Q8DWU5; 306 AA.
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Transcriptional regulator, Cro/CI family.
GN SAG2118.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouiri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014287; AAN00976.1; -
DR TIGR; SAG2118; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF01381; HTH_3; 1.
DR SMART; SM00530; HTH_XRE; 1.
KW Complete proteome.
SQ SEQUENCE 306 AA; 35464 MW; 6B9E7E08B7BDEAFA CRC64;

Query Match 7.9%; Score 84; DB 16; Length 306;
Best Local Similarity 23.8%; Pred. No. 76;
Matches 34; Conservative 25; Mismatches 62; Indels 22; Gaps 6;


```
QY 18 DPHYDKITEINKAI--DDAIAAEQSETIDPMKVPDHPADKFERHVGI--VDFKGLAMRN 75
Db 146 DDIFFMCELTKNISRQEAQTVLSQOEGLDTKIKELSNYPANQIYISYDFVGNL--EN 203
QY 76 IEARGHKKMKRQGDANVKEEGIVKAHLIGVHDDIVSMEYDLAYKLG-----DLHPT 128
Db 204 LSKDKTKQEVKE-----SYERLLKDSPOAG-----VVSYSNVLGLFLINTLADSOPT 252
QY 129 THVISIDIQFVVALSLEISDEGN 151
Db 253 RQLSDMSQATVGMPSQFHEGH 275

RESULT 83
Q9WZQ7 PRELIMINARY; PRT; 314 AA.
AC QSWZQ7;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein TM0800.
GN Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 399:323-329 (1999).
DR EMBL; AE001747; AAD35882.1; -
DR FNR; A72335; A72335.
DR HSSP; P12268; I530.
DR TIGR; TM0800; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR0041136; 2npnp dioxygen.
DR InterPro; IPR003009; FMN_enzyme.
DR Pfam; PF03060; NPD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 314 AA; 33673 MW; 3321EBCF7647D9B8 CRC64;

Query Match 7.9%; Score 84; DB 16; Length 314;
Best Local Similarity 23.5%; Pred. No. 78;
Matches 48; Conservative 32; Mismatches 76; Indels 48; Gaps 10;

QY 23 DKITEINKAIDDAIAA--IEQSETIDPMKVPDHPADKFERHVGI-----GIVDFKGLA 72
Db 122 DSLARVVERAGADAVIAEGSGHGICEVTTFFVLNVKRSVNPVIAAGGADGRGMAA 181
QY 73 MRNIEARGLKKMKR-----QGDNVKGEGIVKAHLIGVHDDIVSMEYDLAYKGLDHP 127
Db 182 AFALGAQVQMGTRFVASVESDVHPYKEKIVKA-----SIRDVVT-----GAKLG--RP 230
QY 128 T-----THVISIDIQFVVALSLEISD-----EGNITMTSFEVROFANVNHIGSL 172
Db 231 ARVLRTPFARKIQEMEPENPMQAEMLVGLSRRAVVEGDLGRGFMVQSGAGLIDEI--- 287
QY 173 SILDPIFGVLSDLVLTAFQDTEVK 196
Db 288 ---KPKVQIILEDILKE--FKETVEK 307

RESULT 84
```

```
Q99X76 PRELIMINARY; PRT; 392 AA.
AC Q99X76;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phosphotransferase (EC 5.4.2.7).
GN DRM OR SAV0139 OR SA0134 OR MW0113.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
Yanamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AP003358; BAB56301.1; -
DR EMBL; AP003129; BAB41354.1; -
DR EMBL; AP004822; BAB93978.1; -
DR FNR; G89774; G89774.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0008973; F:phosphotransferase activity; IEA.
DR InterPro; IPR006124; Metalloenzyme.
DR Pfam; PF01676; Metalloenzyme; 1.
KW Isomerase; Complete proteome.
SQ SEQUENCE 392 AA; 43795 MW; 6012D71CD9BDB549 CRC64;

Query Match 7.9%; Score 84; DB 16; Length 392;
Best Local Similarity 24.1%; Pred. No. 1e+02;
Matches 45; Conservative 35; Mismatches 85; Indels 22; Gaps 10;

QY 14 AVSADPHYDKITE-EINKAIDDAIAAEQSETIDPMKVPDHPADKFERHVGIVDFKGLA 72
Db 65 AVEQPEAYTKLSASVCKDTMTGHWIEMGLNINQPKVYPNGPPELIQOIEMTGRKV 124
QY 73 MRNIEARGLKKMKRQGDANVKEEGIVKAH-----LLIGVHDDIVSME--YDLAYKGLD- 125
Db 125 VANKPASQTQIIDEWGEHOMKTGLIVVTSADPVLQIAAHEDIIPBELYDICEKVELT 184
QY 126 HPTHTVISIDIQFVVALSLEISDEGNITMTS-----FEVROFA-NVNVHI--GGLSILDP 178
Db 185 KDPKYLIGRI-----IARPY-VGEFGNFTRTSNRHDYALKPFGKTVLQHLKDDGGYDVI--A 237
QY 179 FGVLSDV 185
Db 238 IGKINDI 244

RESULT 85
Q9X9I6 PRELIMINARY; PRT; 600 AA.
ID Q9X9I6
```

AC Q9X9I6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Inner membrane ABC-transporter, Irp7.
 GN IRP7.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99426823; PubMed=10496905;
 RA Rakin A., Noelting C., Schubert S., Heesemann J.;
 RT "Common and specific characteristics of the high-pathogenicity island
 of Yersinia enterocolitica.";
 RL Infect. Immun. 67:5265-5274(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99426823; PubMed=10496905;
 RA Rakin A., Schubert S., Gullvout I., Carniel E., Heesemann J.;
 RT "Local hopping of IS3 elements into the A-T-rich part of the high-
 pathogenicity island in Yersinia enterocolitica 1B, O:8.";
 RL FEMS Microbiol. Lett. 182:2255-2294(2000).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AJ132668; CAB46572.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC TM transp.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00664; ABC membrane; 1.
 DR Pfam; PF00005; ABC trans; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 600 AA; 66364 MW; F6C6411A26737EA4 CRC64;
 Query Match 7.9%; Score 84; DB 2; Length 600;
 Best Local Similarity 22.2%; Pred. No. 1.8e+02;
 Matches 43; Conservative 25; Mismatches 66; Indels 60; Gaps 6;
 QY 6 LIAAVAFVAVSADP-----IHYDKITEINKAIDDAIAAIOSETIDPMKVPDHDADKPERH 61
 DB 287 LIAAVAMIRFAEPWAMFISTVSVE-----LIASALQRIEQFVAIAPLPVAAQSEMPERY 342
 QY 62 VGVDFVKGELAMRNIEARGLKQMKRGDANVKGEIGVKAHLIGVHDDIVSVEYDLAYK 121
 DB 343 -----DIRFDNVSYR-----YE 354
 QY 122 LGDLHPHTHVISDIQDFVALSLSEISDEGNITMTSPFVROFANVFN--HIGGLSLIDPI 178
 DB 355 EGDGHALNHLSTLTPAASMSALVAGSAGKTTVKLLMR-YAPQOQISIGGVDIRLRT 413
 QY 179 FGVLSDVLTAIFQD 192
 DB 414 PEQLNLSLVVVFQD 427
 RESULT 86
 Q92375 PRELIMINARY; PRT; 600 AA.
 ID Q92375
 AC Q92375;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE YETQ protein (Inner membrane ABC-transporter YbtQ) (Permease and
 DE ATP-binding protein of yersiniabactin-iron ABC transporter).
 GN YBTQ OR YPO1914 OR Y2396.

OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=6/69;
 RA Buchrieser C., Rusniok C., Couve E., Frangeul L., Billault A.,
 RA Kunst F., Carniel E., Glaser P.;
 RT "DNA sequence of the 102 kbases unstable region of Yersinia pestis.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=KIM6;
 RA MEDLINE=99035519; PubMed=9818149;
 RA Gehring A.M., DeMoll E., Fetherston J.D., Mori I., Mayhew G.F.,
 RA Blattner F.R., Walsh C.T., Perry R.D.;
 RT "Iron acquisition in plague: modular logic in enzymatic biogenesis of
 yersiniabactin by Yersinia pestis.";
 RL Chem. Biol. 5:573-586(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holroyd S., Jageis K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AL031866; CAA21387.1; -.
 DR EMBL; AF091251; AAC69584.1; -.
 DR EMBL; AJ414150; CAC90730.1; -.
 DR EMBL; AE013842; AAM85954.1; -.
 DR FIR; AF0233; AF0233.
 DR FIR; T17436; T17436.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001140; ABC TM transp.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00664; ABC membrane; 1.
 DR Pfam; PF00005; ABC trans; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
 KW ATP-binding; Transport; Complete proteome.
 SQ SEQUENCE 600 AA; 66464 MW; 759638E8250E559C CRC64;
 Query Match 7.9%; Score 84; DB 16; Length 600;
 Best Local Similarity 21.6%; Pred. No. 1.8e+02;
 Matches 42; Conservative 27; Mismatches 65; Indels 60; Gaps 6;
 QY 6 LIAAVAFVAVSADP-----IHYDKITEINKAIDDAIAAIOSETIDPMKVPDHDADKPERH 61
 DB 287 LIAAVAMIRFAEPWAMFISTVSVE-----LIASALQRIEQFVAIAPLPVAAQSEMPERY 342

```
QY 62 VGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEEGIVKAHLIIQVHDDIVSMEYDLAYK 121
Db 343 -----DIRFNVYR-----YE 354
QY 122 LGLHPTHTVISOQFVVALSLEISDEGNITWTSFEVQFANVNV---HIGGLSILDP 178
Db 355 EGDGHALNHVSLTPFAASMSALVGCAGKTTVTKLMLR-YADPQOGQISIGGVDIRLUT 413
QY 179 FGVLSVLTALFOD 192
Db 414 PEOLNLSISVVFOD 427

RESULT 87
Q9CHZ3 PRELIMINARY; PRT; 775 AA.
AC Q9CHZ3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Cation-transporting ATPase.
GN YFGQ OR LL0575.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
CX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006291; AAK04673.1; -.
DR PIR; G86696; G86696.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0006812; P:cation transport; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001757; ATPase_E1-E2.
DR InterPro; IPR008250; E1-E2_ATPase_reg.
DR InterPro; IPR005834; Hydrolyase.
DR Pfam; PF00122; E1-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolyase; 1.
DR PRINTS; PR00119; CATATPASE.
DR TIGRFAMs; TIGR01494; ATPase_P-type; 4.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
KW Complete proteome.
SQ SEQUENCE 775 AA; 85467 MW; 84D5E4E49A833BE CRC64;

Query Match 7.9%; Score 84; DB 16; Length 775;
Best Local Similarity 19.8%; Pred. No. 2.6e+02;
Matches 50; Conservative 37; Mismatches 103; Indels 62; Gaps 6;

QY 22 YDKITEINKAIDATAAIEQSEITDPMKVPDHDKFEHVGIVDFKGL-----AMEN 75
Db 397 YEKVSSINHLLEGFRVLVAGTKE--KIYDQNLGVALGVVLIANPIRENAKSTFNY 454
QY 76 IEARGLKQMKRQGD-----ANVKGEEGIVKAHLI----- 104
Db 455 FAEGGVNKKVIGSDNPQTVSAVAKRAGITGAERFIDANLLKTKEDLDQAVESYVFGRYT 514
QY 105 -----IGVHDDIVSMEYD-----LAYKGLDHPHTHVISDIDQFVALSLEISDE 149
Db 515 PDQKRRLVQALKRKHHTVAMTGVNDILAMKADCSIAWASGSDAATQVAVQLDSD 574
QY 150 GNTITWTSFEVQFANVNVHIGGLSILDPFGLVSDVLTAFQDVTAKEMTK-----V 201
Db 575 GHMTQVTEGRRVNVNVRSAIFLVKNLPSIILAIISALFVFTYPLQASQLSLISLFTI 634
```

```
QY 202 LAPAFKRELEKN 213
Db 635 GIPGFLLSLEEN 646

RESULT 88
Q93NX9 PRELIMINARY; PRT; 9510 AA.
AC Q93NX9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Amph1.
GN AMPH1.
OS Streptomyces nodosus.
OC Bacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., O'Leary M.,
RT "The amphoterin biosynthetic gene cluster from Streptomyces
RT nodosus."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AF357202; AAK73501.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0004334; F:fatty acid biosynthesis; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; AC trans.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR004410; FAD.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR001005; Myb_DNA_Binding.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp bind.
DR Pfam; PF00698; Acyl trans; 6.
DR Pfam; PF00106; adh short; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fadD; 6.
DR PROSITE; PS00075; ACP DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00037; MYB 1; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 6.
KW Oxidoreductase; Phosphopantetheine; Transferase.
SQ SEQUENCE 9510 AA; 992817 MW; 1A8E003A15D478CF CRC64;

Query Match 7.9%; Score 84; DB 2; Length 9510;
Best Local Similarity 24.5%; Pred. No. 7.1e+03;
Matches 53; Conservative 39; Mismatches 80; Indels 44; Gaps 10;

QY 3 KFLIIAFAVAVAGADPHYD---KIT--EINK-----AIDDAIAIEQSE 44
Db 582 RFPFAALDAALDAFTPLHPLKRVLWGEDADRLDRTEVAQPALFAVEVALYLLSEF 641
QY 45 TIDPMKVDPDA--DKFERHGVDFKGLAMRNIEARG-LQMKRQGDANVKGEEGIVKA 101
Db 642 EVKPDHLGSHVGEIAAAHAGV-FSLDDAATLVAARGRLQALPEGAMVAVQASEDEV 700
QY 102 HLLIGVHDDIVSMEYDILAYKGLDHPHTHVISDIDQFVALSLEISDEGNIT----- 153
Db 701 AFLLAGHEDLVY---LAAVNG---PSAVVLGSDETTVELAARLAADGRKTSRLRVSHA 753
QY 154 ---WTSFEVQFANVNVHIGGLSILDPFGLVSDV 185
Db 754 PHSPLMAPMLDEFNRVVE---GLTILSPILPVSVDV 786
```

```

RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J., Maskell J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43727.1; -.
KW Complete proteome.
SQ SEQUENCE 266 AA; 28658 MW; 7207543ADBFF8F6F CRC64;

Query Match 7.8%; Score 83.5; DB 16; Length 266;
Best Local Similarity 24.4%; Pred. No. 69;
Matches 53; Conservative 32; Mismatches 75; Indels 57; Gaps 12;

QY 3 KFLIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-AIAAEQSESTID---PMK 50
Db 17 RYLATAVVALAVAAALYARQDEPEHEDEV---VLLCLDDQAMSTVTRTGSVDGILPAC 73
QY 51 VPDHAKFERHVGIVDPFKGELAMRNIEAR-GLKQMKRQGDANVKGEGIVKAHLI----- 104
Db 74 QPELAELVGEDFGAVD---EPTLRVAVLATVAANFAEYSSSAVACEAIKESQHLNCCNL 130
QY 105 -----IGVHDDIVSMEYDLAYKLGID---LHPTTHVISDIQDFVVAL 142
Db 131 VFLMGVLYGLKEKGIPIGVGGAGVGNHQAQIMTKSGNNRLLDPTTGCVAAI-TFNKLM 189
QY 143 SLEISDEGNI-----TMTSFEVRQFANVNNHIG 171
Db 190 SGQPVPEEQVRVFRVKATIDSFDRVYAVVG--GG 224

RESULT 91
Q93R43 ID Q93R43 PRELIMINARY; PRT; 311 AA.
AC Q93R43;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
GN HSP60.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=630;
RN [1]_TaxID=630;
RP SEQUENCE FROM N.A.
RA Yang R.P.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Du Z.M.;
RT "Bacillus and Versinia HSP60 gene partial sequence.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064591; BAB61892.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1 311
FT NON_TER 311 311
SQ SEQUENCE 311 AA; 33032 MW; 34429B1392166D86 CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 311;
Best Local Similarity 18.1%; Pred. No. 85;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

QY 13 VAVSADPHYDKITEINKAIDDA-IAAEQSESTIDPMKVPDHDHAKFERHVGIVDFKGL 71
Db 16 VAAAGNPNM-----DLKRGIDKAVIAAVEE---LKKLSVPCSDSKAIAQAQVGTISANDS 65
QY 72 AMRNIARGLKQMKRQGDANVKGEGIVKAHLIIGVHDD---IVSMEYDLAY----- 120

RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J., Maskell J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43727.1; -.
KW Complete proteome.
SQ SEQUENCE 266 AA; 28658 MW; 7207543ADBFF8F6F CRC64;

Query Match 7.8%; Score 83.5; DB 16; Length 266;
Best Local Similarity 24.4%; Pred. No. 69;
Matches 53; Conservative 32; Mismatches 75; Indels 57; Gaps 12;

QY 3 KFLIAAVAF-VAVSA-----DPIHYDKITEINKAIDD-AIAAEQSESTID---PMK 50
Db 17 RYLATAVVALAVAAALYARQDEPEHEDEV---VLLCLDDQAMSTVTRTGSVDGILPAC 73
QY 51 VPDHAKFERHVGIVDPFKGELAMRNIEAR-GLKQMKRQGDANVKGEGIVKAHLI----- 104
Db 74 QPELAELVGEDFGAVD---EPTLRVAVLATVAANFAEYSSSAVACEAIKESQHLNCCNL 130
QY 105 -----IGVHDDIVSMEYDLAYKLGID---LHPTTHVISDIQDFVVAL 142
Db 131 VFLMGVLYGLKEKGIPIGVGGAGVGNHQAQIMTKSGNNRLLDPTTGCVAAI-TFNKLM 189
QY 143 SLEISDEGNI-----TMTSFEVRQFANVNNHIG 171
Db 190 SGQPVPEEQVRVFRVKATIDSFDRVYAVVG--GG 224

RESULT 91
Q93R43 ID Q93R43 PRELIMINARY; PRT; 311 AA.
AC Q93R43;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
GN HSP60.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=630;
RN [1]_TaxID=630;
RP SEQUENCE FROM N.A.
RA Yang R.P.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Du Z.M.;
RT "Bacillus and Versinia HSP60 gene partial sequence.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064591; BAB61892.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1 311
FT NON_TER 311 311
SQ SEQUENCE 311 AA; 33032 MW; 34429B1392166D86 CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 186;
Best Local Similarity 24.1%; Pred. No. 43;
Matches 41; Conservative 25; Mismatches 65; Indels 39; Gaps 8;

QY 13 VAVSADPHYDKITEINKAIDDAIAAEQSESTIDPMKVPDHDHAKFERHVGIVDFKGLA 72
Db 16 VTAGASPI-----DLKRGIDKAVAAVVAELKAHSHEVGDDYSKVEQ-VGTVAANDAT 67
QY 73 MRNIARGLKQMKRQGDANV---KGEIGVKAHLIIGVHDDIVSMEYDLAYKLGIDLHPTT 129
Db 68 IGKLIADAMSKVKDGVITVEAKGTETVKV-----VEGMQFDRGV----ISPYF 114
QY 130 HVISDIQDFVVALSLEIS-DEGNTMTSFEVRQFANVNNHIGLSILDPI 178
Db 115 MTNSD-----KMEASLDSPPYLITDKKISTMKDV-----LPFILEPI 150

RESULT 90
Q7VTQ1 ID Q7VTQ1 PRELIMINARY; PRT; 266 AA.
AC Q7VTQ1;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative exported protein.
GN BP3464.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Tohama I / ATCC BAA-589 / NCIC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker I., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
```

```

Db 66 TVGELIAQAMEKV-----KLGDLHPTTHVISDIQDFVVALSLEISDEGNITMTSFEV 116
Qy 121 -----KLGDLHPTTHVISDIQDFVVALSLEISDEGNITMTSFEV 159
Db 117 PETGSIELESPFILLADKKISNIREMPLVLEAVAKAGKPLIIAEDVEGEALATL----- 171
Qy 160 RQFANVNNHIGGL-----SILDPIFGVLSVDLTAIFQDVTVRKEMTKVLAPAPKRELEK 212
Db 172 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEIGLELEK 220

RESULT 92
Q93R44
ID Q93R44 PRELIMINARY; PRT; 311 AA.
AC Q93R44
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).
GN HSP60.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang R.F.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Du Z.M.;
RT "Bacillus and Yersinia HSP60 gene partial sequence.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064590; BAB61891.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:Chaperone activity; IEA.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1 1
FT NON_TER 311 311
SQ SEQUENCE 311 AA; 33002 MW; C1428E0692166D9C CRC64;

Query Match 7.8%; Score 83.5; DB 2; Length 311;
Best Local Similarity 18.1%; Pred. No. 85;
Matches 43; Conservative 45; Mismatches 80; Indels 69; Gaps 9;

Qy 13 VAVSADPIHYDKITEINKAIDDA-TAAIEQSETIDPMKVPDHPADKFERHVGIVDFKGE 71
Db 16 VAGMNP-----DLKRGIDKAVIAAVEE---LKKLSVPCSDSKAIAQVGTISANS 65
Qy 72 AMRNIIEARGLKQMKROGDANVKGEIGIVKAHLIIGVHDD---IVSMEYDLAY----- 120
Db 66 TVGELIAQAMEKV-----KLGDLHPTTHVISDIQDFVVALSLEISDEGNITMTSFEV 159
Qy 121 -----KLGDLHPTTHVISDIQDFVVALSLEISDEGNITMTSFEV 159
Db 117 PETGSIELESPFILLADKKISNIREMPLVLEAVAKAGKPLIIAEDVEGEALATL----- 171
Qy 160 RQFANVNNHIGGL-----SILDPIFGVLSVDLTAIFQDVTVRKEMTKVLAPAPKRELEK 212
Db 172 -----VNTMRGIVKVAAPGFG---DRRKAMLDIATLTAGTVISEIGLELEK 220

RESULT 94
Q93R44
ID Q93R44 PRELIMINARY; PRT; 329 AA.
AC Q93R44
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE groEL-like protein (60 kDa chaperonin) (Protein Cpn60) (groEL protein)
DE (Fragment).
GN GROEL.
OS Enterobacteriaceae sp. JN965.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae.
OX NCBI_TaxID=146477;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=JN965;
RA Rascoc J.B.; Berg M., Bruton B., Pair S., Mitchell F., Melcher U.,
RA Fletcher J.;
RT "Characterization of Serratia isolates associated with cucurbit yellow
vine disease.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

```

```

Q93R42
ID Q93R42 PRELIMINARY; PRT; 311 AA.
AC Q93R42
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HSP60 (Fragment).

```

```

DR EMBL; AJ302157; CAC19389.1; -
DR HSP; P06139; 10RL.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR InterPro; IPR001844; Chapnlin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
KW ATP-binding; Chaperone.
FT NON_TER 329 329
SQ SEQUENCE 329 AA; 34639 MW; F6C7177B5E8B15 CRC64;

Query Match
Best Local Similarity 18.1%; Pred. No. 92;
Matches 43; Conservative 44; Mismatches 81; Indels 69; Gaps 9;

QY 13 VASADPHYDKITEEINKAIDDAI-AAIQSEITIDPMKVPDHDADKFERHVGIVDFKGL 71
Db 107 VAAGNPNP-----DLKRGIDKAVTAABE---LKALSVPSCDSKAIAQVGISANSDE 156
QY 72 AMRIEARGLKKOMKQGDANVKGEGIVKAHLIGVHDD---IVSMYEDLAY----- 120
Db 157 TVGKLIAEMEKV-----GREGVITVEEGTGLQDELVDVWEGMDFDRGYLSPYFINK 207
QY 121 -----KLGDLHPTTHVDSIQDFVVALSLEISDEGNITMTSFEV 159
Db 208 PETGSVELESFILLADKKISNIREMLFVLEAVAKAGKPLVIAEDVEGALATL----- 262
QY 160 RQFANVNHGGL---SILDPIFGVLSDLVTAIFQDTVRKMTKVLAPAFKRELEK 212
Db 263 -----VNTMRGIVKVAVKAPGFG---DRKAMLDIATITGCTVISEIHCMELEK 311

RESULT 95
Q8K3U8 PRELIMINARY; PRT; 402 AA.
AC Q8K3U8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE FK506 binding protein 4 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Desai B.J., McKinney K.Q., Meyer M.H., Bahrani-Mostafavi Z.,
RA Meyer R.A. Jr.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF531427; AAM95632.1; -
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF00254; FKBP; 2.
DR Pfam; PF00315; TPR; 3.
DR SMART; SM00026; TPR; 3.
DR PROSITE; PS00454; FKBP_PPIASE_2; 2.
DR PROSITE; PS00059; FKBP_PPIASE_3; 2.
FT NON_TER 1
SQ SEQUENCE 402 AA; 45486 MW; 82D74DB9643D1B51 CRC64;

Query Match
Best Local Similarity 22.1%; Pred. No. 1.2e+02;
Matches 27; Conservative 30; Mismatches 48; Indels 17; Gaps 3;

QY 18 DPHYDKITEEINKAIDDAIAIQSEITI-----DPMKVPDHDAD-KFERHV 62
Db 19 DRFSFDLKGKIVKAWDIAVATMKVGVCHITCKPEYAGSGSPKPPNATLVFE--V 76

EMBL; AJ302157; CAC19389.1; -
77 ELFEFKGDELDEDEGGIIRTRGEGYARPDGAMVEVALEGYINDRLFDQRELCEV 136
123 GD 124
137 GE 138

RESULT 96
Q8NUB6 PRELIMINARY; PRT; 412 AA.
AC Q8NUB6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 5'-nucleotidase/2',3'-cyclic phosphodiesterase and related
DE esterases.
GN CGL0023.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=17118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005274; BAB97416.1; -
DR GO; GO:0016787; P:hydrolase activity; IEA.
DR GO; GO:0009166; P:nucleotide catabolism; IEA.
DR InterPro; IPR008334; 5'-Nucleotidase C.
DR InterPro; IPR006179; 5-nucleotidase.
DR InterPro; IPR001064; Cystallin.
DR Pfam; PF02872; 5_nucleotidaseC; 1.
DR PRINTS; PR01607; APYRASEFAMILY.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
KW Complete proteome.
SQ SEQUENCE 412 AA; 43019 MW; C48CA12D657A5B02 CRC64;

Query Match
Best Local Similarity 31.5%; Pred. No. 1.2e+02;
Matches 34; Conservative 16; Mismatches 41; Indels 17; Gaps 6;

QY 54 HADKFERHVGIVDFKXGELAMRNIEARGLKKOMKQGDANVKGEGIVKAHLIGVHDDIVS 113
Db 134 YAEAFE---IQFSGSDSFVTLKGDFKDALEQ-----QWEEGSAFPVAALGVSDN-VS 183
QY 114 MEYDLAYKLGDLHPHTHTVDSI-----QDFVVALSLEISDEGNITMTS 156
Db 184 YTYDINRPIGD-RVTSVTIDTPTLPDPERDYVVAASLYL-QSGNEGTA 229

RESULT 97
Q97K37 PRELIMINARY; PRT; 474 AA.
AC Q97K37;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Beta-glucosidase family protein.
GN CAC1084.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,

```

Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
Ratsov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
J. Bacteriol. 183:4823-4838(2001).
EMBL; AB007625; AAK79058.1; -.
DR PFAM; G97033; G97033.
DR GO; GO:0004553; F-hydrolase activity, hydrolyzing O-glycosyl . . ; IEA.
GO; GO:0005975; P-carbonhydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PR00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD00650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1;
PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 55440 MW; 4F34109B7735168B CRC64;

Query Match 7.8%; Score 83.5; DB 16; Length 474;
Best Local Similarity 23.3%; Pred No. 1.5e+02;
Matches 34; Conservative 22; Mismatches 65; Indels 25; Gaps 3;

DG 53 DHADKPERHVGIVDFKGELAMNIEARGLKQMKRGQGDANVKGE---GIWKAHLIGVHD 109
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
55 DIVHFREKDVALMAEGLKAYRFVSWSKRIPYGRGVNLKGLEFYKNLDLKYNIENP 114
||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
110 DIVSMEDYLAIKGLDH---PTTHVISDIQDFVFALSLEISDEGNITWTSEVRQFAN-- 164
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
115 VLTLHWDLPLQLDYGFERNVISDFENICITLFNFNDKVYKITFNQNVTNLG 174
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
165 -----VNHHGGJS 173
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
175 YRCVHPPIKDKIKYYLANHVLA 200

RESULT 98
QCUB4 PRELIMINARY; PRT; 1047 AA.

ID QCUB4 AC
AC Q9CQB4
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein KIAA1749 (Fragment).
GN KIAA1749.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
[1] _SEQUENCE FROM N.A.
RP Tissue=Brain;
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okamura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RNA Res. 7:347-355(2000).
RL EMBL; AB051536; BAB21840.2; -.
DR InterPro; IPR002928; Myosin tail.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PR01576; Myosin tail; 1.
DR PRINTS; PR00194; TROPOMYSIN.
FW Hypothetical protein.
KT NON_TER
FT SEQUENCE 1047 AA; 120716 MW; 2F1FACA1357A79CF CRC64;

Query Match 7.8%; Score 83.5; DB 4; Length 1047;
Best Local Similarity 21.7%; Pred. No. 4.2e+02;
Matches 31; Conservative 29; Mismatches 56; Indels 17; Gaps 4;

DG 21 HYDKITEINKAIDAIAEQSETIDPMKVPDHADKFERHVGIVFKGELAMNTEAR- 79
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

[illegible]

Mon Aug 9 08:14:05 2004

DR Pfam; PF00118; cpn60_TCP1; 1.
KW ATP-binding; Chaperone.
FT NON_TER 1
FT NON_TER 186 186
SQ SEQUENCE 186 AA; 20213 MW; E5919F6D83F16A0E CRC64;

Query Match 7.8%; Score 83; DB 2; Length 186;
Best Local Similarity 23.3%; Pred.No. 47;
Matches 27; Conservative 26; Mismatches 49; Indels 14; Gaps 4;

Qy 5 LLIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAIEQSETIDPMKVPDHDADKFERHVGI 64
Db 8 ILTEGMKVAAAGAPL-----DVKRGIDKAVAKVVEIKSQAEIYGDYDKIEQ-VGT 59

Qy 65 VDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120
Db 60 VSANNDPVIGKLIADAMRKVKSGDGVITIEAKG---TDTTIGV---VEGMQFDRGY 109

Search completed: August 6, 2004, 16:01:44
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 16:02:17 ; Search time 54 Seconds
(without alignments)
1114.493 Million cell updates/sec

Title: US-10-024-955-7

Perfect score: 213

Sequence: 1 MMKFLIIAAVAFVAVSADPI.....VRKEMTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A_Geneseq_29Jan04:**

- 1: Geneseqp1980s:**
- 2: Geneseqp1990s:**
- 3: Geneseqp2000s:**
- 4: Geneseqp2001s:**
- 5: Geneseqp2002s:**
- 6: Geneseqp2003as:**
- 7: Geneseqp2003bs:**
- 8: Geneseqp2004s:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	100.0	213	2	AAR60576
2	213	100.0	213	2	AAY25592
3	213	100.0	213	7	ADC3482
4	25	11.7	215	2	AAR60575
5	25	11.7	215	2	AAY25586
6	25	11.7	215	7	ADC34836
7	8	3.8	87	5	ABP27647
8	8	3.8	321	3	AAG31332
9	8	3.8	342	3	AAG31331
10	8	3.8	349	3	AAG31330
11	8	3.8	412	3	AAR41823
12	8	3.8	484	6	ABU23355
13	8	3.8	580	2	AAW68402
14	8	3.8	611	4	AAM50122
15	8	3.8	783	5	ABBS2052
16	8	3.8	1784	2	AAR05898
17	8	3.8	1784	2	AAR34427
18	7	3.3	73	3	AAB51943
19	7	3.3	85	6	ABR41325
20	7	3.3	87	4	AAU37687
21	7	3.3	87	5	ABP27648
22	7	3.3	87	6	ABU01601
23	7	3.3	87	6	ABU46065
24	7	3.3	87	6	ABU44650
25	7	3.3	87	6	ABU46703

26	7	3.3	133	3	AAG01133	Aag01133 Human sec
27	7	3.3	139	4	AAU30907	Aau30907 Novel hum
28	7	3.3	148	6	ADA33025	Ada33025 Acinetoba
29	7	3.3	152	4	AAU30909	Aau30909 Novel hum
30	7	3.3	155	3	AAG61764	Aag61764 Arabidops
31	7	3.3	178	5	ABP26815	Abp26815 Streptoco
32	7	3.3	178	5	ABP26815	Abp26815 Streptoco
33	7	3.3	178	5	ABP26815	Abp26815 Streptoco
34	7	3.3	181	3	AAG59755	Aag59755 Arabidops
35	7	3.3	183	3	AAG59754	Aag59754 Arabidops
36	7	3.3	183	3	AAG59754	Aag59754 Arabidops
37	7	3.3	185	3	AAG55324	Aag55324 Arabidops
38	7	3.3	190	5	ABP69706	Abp69706 Human pol
39	7	3.3	192	7	ADA24313	Ada24313 Bacillus
40	7	3.3	193	5	ABB06132	Abb06132 Human NS
41	7	3.3	204	2	AAR54114	Aar54114 GDI D4 pr
42	7	3.3	205	3	AAG59753	Aag59753 Arabidops
43	7	3.3	214	3	AAG55323	Aag55323 Arabidops
44	7	3.3	215	3	AAG55906	Aag55906 Arabidops
45	7	3.3	267	7	ADB64094	Adb64094 Human pro
46	7	3.3	295	7	ADB85214	Adb85214 Rat chima
47	7	3.3	300	3	AAB18216	Aab18216 Plasmodiu
48	7	3.3	343	6	ADB09304	Adb09304 Alloiococ
49	7	3.3	354	2	AAR33439	Aar33439 Ornithine
50	7	3.3	368	6	ABU35415	Abu35415 Protein e
51	7	3.3	372	6	ABU40439	Abu40439 Protein e
52	7	3.3	383	5	ABP69823	Abp69823 Human pol
53	7	3.3	425	6	ADB09306	Adb09306 Alloiococ
54	7	3.3	430	6	ABU23589	Abu23589 Protein e
55	7	3.3	434	5	ABB54464	Abb54464 Lactococc
56	7	3.3	453	6	ABU24395	Abu24395 Protein e
57	7	3.3	460	4	ABB63535	Abb63535 Drosophil
58	7	3.3	466	6	ABR58660	AbR58660 Human can
59	7	3.3	468	6	ABU25000	Abu25000 Protein e
60	7	3.3	468	7	ADE59897	Ade59897 Human pro
61	7	3.3	474	5	ABB2783	Abb2783 Herbicida
62	7	3.3	476	6	ADB09308	Adb09308 Alloiococ
63	7	3.3	498	6	ABP71433	Abp71433 Y. peatis
64	7	3.3	511	4	ABB62629	Abb62629 Drosophil
65	7	3.3	521	6	ADB09310	Adb09310 Alloiococ
66	7	3.3	535	6	ABJ39024	Abj39024 Mouse cla
67	7	3.3	535	7	ADC17327	Adc17327 Mouse IFN
68	7	3.3	581	4	ABG35418	Abg35418 Novel hum
69	7	3.3	581	5	ABG97496	Abg97496 Human NOV
70	7	3.3	625	6	ABU20412	Abu20412 Protein e
71	7	3.3	657	7	ADB80157	Adb80157 Mycobacte
72	7	3.3	752	2	AAR38153	Aar38153 Acetobact
73	7	3.3	791	5	ABB93713	Abb93713 Herbicida
74	7	3.3	887	6	ADB11976	Adb11976 Alloiococ
75	7	3.3	1001	6	ABU21679	Abu21679 Protein e
76	7	3.3	1027	7	ADC01332	Adc01332 Enteroha
77	7	3.3	1194	6	ABM15822	Abm15822 Mycobacte
78	7	3.3	1400	7	ADC01365	Adc01365 Enteroha
79	7	3.3	2000	6	ABR52951	AbR52951 Protein s
80	7	3.3	2836	4	ABB62719	Abb62719 Drosophil
81	6	2.8	8	4	AAB23790	Aab23790 Voltage s
82	6	2.8	8	5	ABG70880	Abg70880 R. rhodoc
83	6	2.8	10	2	AAR70305	Aar70305 Laminin (
84	6	2.8	10	5	ABG67586	Abg67586 Human ADP
85	6	2.8	10	6	ABU58121	Abu58121 Laminin c
86	6	2.8	10	6	ADA23705	Ada23705 Alzheimer
87	6	2.8	11	2	AAW69442	Aaw69442 Rat phosph
88	6	2.8	15	2	AAW73718	Aaw73718 M. tuberc
89	6	2.8	15	2	AAW73716	Aaw73716 M. tuberc
90	6	2.8	15	2	AAW73826	Aaw73826 M. tuberc
91	6	2.8	15	2	AAW73828	Aaw73828 M. tuberc
92	6	2.8	15	4	AAU08183	Aau08183 Mycobacte
93	6	2.8	15	4	AAU08181	Aau08181 Mycobacte
94	6	2.8	19	2	AAR25844	Aar25844 Laminin-p
95	6	2.8	19	2	AAR44034	Aar44034 Laminin-d
96	6	2.8	19	2	AAR92741	Aar92741 IKVAV-con
97	6	2.8	19	2	AAI22410	Aay22410 Laminin a
98	6	2.8	19	2	AAI31007	Aay31007 Non-cross

99 6 2.8 19 5 AA48193 Peptide #
100 6 2.8 19 6 AB090762 Peptide #

ALIGNMENTS

RESULT 1

AA060576
ID AA060576 standard; protein, 213 AA.

XX AC AA060576;

DT 25-MAR-2003 (revised)

DT 01-APR-1995 (first entry)

XX DE House dust mite allergen DerfVII.

XX KW DerfVII allergen; antiallergic; allergy diagnosis.

XX OS Dermatophagoides farinae.

XX PN WO9420614-A1.

XX PD 15-SEP-1994.

XX PF 11-MAR-1994; 94WO-AU000117.

XX PR 12-MAR-1993; 93US-00031141.

XX PR 22-JUN-1993; 93US-00081540.

XX PA (CHIL-) INST CHILD HEALTH RES.

XX PI Thomas WR, Chua K;

XX DR WPI; 1994-303021/37.

XX DR N-PSDB; AAQ71401.

XX PT New nucleic acid encoding specific dust mite allergens - and related
PT vectors, transformed cells, peptides and antibodies, useful for
PT desensitisation and diagnosis.

XX PS Claim 5; Page 40-41; 67pp; English.

XX CC DerfVII antigen is useful as antiallergic reagent for treating

XX CC sensitivity to house dust mite allergens. (updated on 25-MAR-2003 to

XX CC correct PN field.)

XX SQ Sequence 213 AA;

Query Match 100.0%; Score 213; DB 2; Length 213;

Best Local Similarity 100.0%; Pred. No. 1.3e-203; Mismatches 0; Indels 0; Gaps 0;

Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAEQSETIDPMKVPDADKFER 60

DB 1 MMKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAEQSETIDPMKVPDADKFER 60

QY 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120

DB 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120

QY 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPFG 180

DB 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPFG 180

QY 181 VLSVLTATFQDVTVRKEMTKVLAPAFKRELEKN 213

DB 181 VLSVLTATFQDVTVRKEMTKVLAPAFKRELEKN 213

RESULT 2

AA25592

ID AA25592 standard; protein, 213 AA.

XX AC AA25592;

DT 30-SEP-1999 (first entry)

XX DE D. farinae allergen Der f 7 protein fragment.

XX KW Major histocompatibility complex; class II; desensitising; human;

XX KW allergen; grass; tree; weed; mite; housefly; fruit fly; sheep blow fly; honeybee;

XX KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;

XX KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;

XX KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;

XX KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX OS Dermatophagoides farinae.

XX PN WO9934826-A1.

XX PD 15-JUL-1999.

XX PF 11-JAN-1999; 99WO-CB000080.

XX PR 09-JAN-1998; 98GB-00000445.

XX PR 21-SEP-1998; 98GB-00020474.

XX PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX PI Larche M, Kay AB;

XX DR WPI; 1999-458255/38.

XX PT Desensitizing patients to polypeptide allergens.

XX PS Example 6; Page 52; 117pp; English.

XX CC This invention describes a novel method of desensitizing a patient to a
XX CC polypeptide allergen and comprises administering to the patient a peptide
XX CC derived from the allergen where restriction to a MHC Class II molecule
XX CC possessed by the patient can be demonstrated for the peptide and the
XX CC peptide is able to induce a late phase response in an individual who
XX CC possesses the MHC Class II molecule. The methods can be used for
XX CC desensitizing patients to allergens present in e.g. grass, tree and weed
XX CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
XX CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit
XX CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
XX CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
XX CC tenbriol molitor beetle, mammals such as cat, dog, horse, cow, pig,
XX CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
XX CC produce immunological vaccines which may be used to prevent and/or treat
XX CC conditions involving hypersensitivity to allergens. This sequence
XX CC represents the Dermatophagoides farinae allergen Der f 7

XX SQ Sequence 213 AA;

Query Match 100.0%; Score 213; DB 2; Length 213;

Best Local Similarity 100.0%; Pred. No. 1.3e-203;

Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAEQSETIDPMKVPDADKFER 60

DB 1 MMKFLIIAAVAFVAVSADPIHYDKITEEINKAIDDAIAAEQSETIDPMKVPDADKFER 60

QY 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120

DB 61 HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEIGVKAHLIGVHDDIVSMEYDLAY 120

QY 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPFG 180

DB 121 KLGLDLPHTTHVISDIQDFVWALSLEISDEGNITMTSFEVROFANVNVHIGLSILDPFG 180

QY 181 VLSVLTATFQDVTVRKEMTKVLAPAFKRELEKN 213

```
Db 181 VLSDLVLTAFQDITVRKEMTKVLAPAKFKELEKN 213

RESULT 3
ADC34842
ID ADC34842 standard; protein; 213 AA.
XX AC ADC34842;
XX AC ADC34842;
DT 18-DEC-2003 (first entry)
XX DE
XX DE House dust mite allergen Der f 7.
XX KW house dust mite; allergen; antigen; hyporesponsive; desensitisation;
XX KW immunomodulator; gene therapy.
XX OS
XX OS Dermatophagoides farinae.
XX PN WO2003047618-A2.
XX PN
XX PD 12-JUN-2003.
XX PF 05-DEC-2002; 2002WO-GB005548.
XX PR 05-DEC-2001; 2001US-0338385P.
XX PR (CIRC-) CIRCASSIA LTD.
XX PA
XX PI Larche M, Ledger PW;
XX PI WPI; 2003-523267/49.
XX PS
XX PS Desensitizing an individual to a selected polypeptide antigen comprises
XX PT administering a composition containing polypeptide antigens in an amount
XX PT that generates a state of hyporesponsiveness to the antigen to allow
XX PT desensitization.
XX PT
XX SQ Disclosure; Page 23; 57pp; English.

Query Match 100.0%; Score 213; DB 7; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e-203; Mismatches 0; Indels 0; Gaps 0;
Matches 213; Conservative 0;

QY 1 MMKFLIIAFAVAVSADPIHYDKITEEINKAIDDAIAAIEQSEITDPMKVPDHADKFER 60
DB 1 MMKFLIIAFAVAVSADPIHYDKITEEINKAIDDAIAAIEQSEITDPMKVPDHADKFER 60
QY 61 HVGIVDFKGLAWRNTEARGLKQKQGDANVKGEIVKAHLICVHDDIVSMYDLAY 120
DB 61 HVGIVDFKGLAWRNTEARGLKQKQGDANVKGEIVKAHLICVHDDIVSMYDLAY 120
QY 121 KLGDLPHTTHVTSIDIQDFVALLSLEISDEGNITMTSFEVRQFANVNVNHIGLSILDPFG 180
DB 121 KLGDLPHTTHVTSIDIQDFVALLSLEISDEGNITMTSFEVRQFANVNVNHIGLSILDPFG 180
QY 181 VLSDLVLTAFQDITVRKEMTKVLAPAKFKELEKN 213
DB 181 VLSDLVLTAFQDITVRKEMTKVLAPAKFKELEKN 213

us-10-024-955-7.oligo.rag

RESULT 4
AAR60575
ID AAR60575 standard; protein; 215 AA.
XX AC AAR60575;
XX AC AAR60575;
DT 25-MAR-2003 (revised)
DT 01-APR-1995 (first entry)
XX DE
XX DE House dust mite allergen DerpVII cDNA.
XX KW DerpVII allergen; antiallergic; allergy diagnosis.
XX OS
XX OS Dermatophagoides pteronyssinus.
XX PN WO9420614-A1.
XX PN
XX PD 15-SEP-1994.
XX PF 11-MAR-1994; 94WO-AU000117.
XX PR 12-MAR-1993; 93US-00031141.
XX PR 22-JUN-1993; 93US-00081540.
XX PR (CHIL-) INST CHILD HEALTH RES.
XX PA
XX PI Thomas WR, Chua K;
XX PI WPI; 1994-303021/37.
XX DR N-PSDB; AAQ71400.
XX CC New nucleic acid encoding specific dust mite allergens - and related
XX PT vectors, transformed cells, peptides and antibodies, useful for
XX PT desensitisation and diagnosis.
XX PT
XX PS Claim 7; Page 36-37; 67pp; English.
XX CC DerpVII antigen is useful as antiallergic reagent for treating
XX CC sensitivity to house dust mite allergens. The DNA can be used as a probe
XX CC to detect the sensitivity of an individual to the allergen. (Updated on 25
XX CC -MAR-2003 to correct PN field.)
XX SQ Sequence 215 AA;
Query Match 11.7%; Score 25; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.2e-16; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

QY 155 TSFEVROFANVNVNHIGLSILDPF 179
DB 155 TSFEVROFANVNVNHIGLSILDPF 179

RESULT 5
AAY25586
ID AAY25586 standard; protein; 215 AA.
XX AC AAY25586;
XX AC AAY25586;
DT 30-SEP-1999 (first entry)
XX DE
XX DE D. pteronyssinus allergen Der p 7 protein fragment.
XX KW Major histocompatibility complex; class II; desensitising; human;
XX KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;
XX KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
XX KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
XX KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
XX KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX OS Dermatophagoides pteronyssinus.
```


PT detecting a compound that binds to the protein.
XX Claim 1; Page 3608; 4525pp; English.
PS
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 87 AA;
Query Match 3.8%; Score 8; DB 5; Length 87;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 DDAAIAIE 41
Db 71 DDAAIAIE 78
|||||||
RESULT 8
AAG31332
ID AAG31332 standard; protein; 321 AA.
XX AAG31332;
AC AAG31332;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37610.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135829P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.

PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142320P.
PR 12-JUL-1999; 99US-0142377P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144322P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148584P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 07-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158023P.
PR 12-OCT-1999; 99US-0158232P.
PR 13-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159373P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

```
PR 25-OCT-1999; 99US-0161405P.
PR 26-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.8%; Score 8; DB 3; Length 342;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QDFVVALS 143
DB 325 QDFVVALS 332

RESULT 10
AAG31330
ID AAG31330 standard; protein; 349 AA.
XX
AC AAG31330;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37608.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; Gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144844P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
```



```
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P.
PR 10-AUG-1999; 99US-0148319P.
PR 11-AUG-1999; 99US-0148341P.
PR 12-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148568P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 31-AUG-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 01-SEP-1999; 99US-0153070P.
PR 10-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154779P.
PR 20-SEP-1999; 99US-0155139P.
PR 22-SEP-1999; 99US-0155486P.
PR 23-SEP-1999; 99US-0155859P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 29-SEP-1999; 99US-0157117P.
PR 04-OCT-1999; 99US-0157753P.
PR 05-OCT-1999; 99US-0157965P.
PR 06-OCT-1999; 99US-0158029P.
PR 07-OCT-1999; 99US-0158232P.
PR 08-OCT-1999; 99US-0158369P.
PR 12-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160742P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.

99US-0147038P.
99US-0147204P.
99US-0147302P.
99US-0147192P.
99US-0147260P.
99US-0147303P.
99US-0147416P.
99US-0147493P.
99US-0147935P.
99US-0148171P.
99US-0148319P.
99US-0148341P.
99US-0148565P.
99US-0148568P.
99US-0149368P.
99US-0149175P.
99US-0149426P.
99US-0149722P.
99US-0149723P.
99US-0149929P.
99US-0149902P.
99US-0149930P.
99US-0150566P.
99US-0150884P.
99US-0151065P.
99US-0151066P.
99US-0151080P.
99US-0151303P.
99US-0151438P.
99US-0151930P.
99US-0152363P.
99US-0153070P.
99US-0153758P.
99US-0154018P.
99US-0154039P.
99US-0154779P.
99US-0155139P.
99US-0155486P.
99US-0155859P.
99US-0156458P.
99US-0156596P.
99US-0157117P.
99US-0157753P.
99US-0157965P.
99US-0158029P.
99US-0158232P.
99US-0158369P.
99US-0159293P.
99US-0159294P.
99US-0159295P.
99US-0159329P.
99US-0159330P.
99US-0159331P.
99US-0159637P.
99US-0159638P.
99US-0159584P.
99US-0160741P.
99US-0160742P.
99US-0160788P.
99US-0160770P.
99US-0160814P.
99US-0160815P.
99US-0160980P.
99US-0160981P.
99US-0160989P.
99US-0161404P.
99US-0161405P.
99US-0161406P.
99US-0161359P.
99US-0161360P.
99US-0161361P.
99US-0161920P.
99US-0161922P.

PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.8%; Score 8; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QDFVVALS 143
DB 332 QDFVVALS 339
|||||
|||||

RESULT 11
AAR41823
ID AAR41823 standard; protein; 412 AA.
XX AAR41823;
XX AAR41823;
XX 24-OCT-2003 (revised)
DT 23-MAR-1994 (first entry)
XX Methylobacillus glycohenes 1006 homoserine dehydrogenase.
DE L-threonine biosynthesis; homoserine dehydrogenase; threonine synthase;
KW homoserine kinase; amino acid; ATCC 21371.
XX Methylobacillus glycohenes; strain 1006 (ATCC 21371).
OS JP05207886-A.
XX 20-AUG-1993.
XX 29-JAN-1992; 92JP-00014335.
XX 29-JAN-1992; 92JP-00014335.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX WPI: 1993-297465/38.
DR N-PSDB; AAQ48930.
XX Prepn. of L-threonine by fermentation in methanol soln. - using
PT Methylobacillus sp. contg. recombinant vector to produce homoserine-
PT dehydrogenase or kinase, or threonine synthase.
XX Claim 5; Page 14-15; 21pp; Japanese.
XX DNA coding for at least one of the enzymes homoserine dehydrogenase,
CC threonine synthase and homoserine kinase is incorporated into a
CC recombinant expression vector and used to transform Methylobacillus
CC bacteria. The transformants are cultured in medium containing methanol as
CC carbon source to synthesise L-threonine. See also AAQ48929 and AAQ48931-
CC Q48932. (Updated on 24-OCT-2003 to standardise OS field)
XX Sequence 412 AA;
SQ
Query Match 3.8%; Score 8; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDAAIAIE 41
DB 386 DDAAIAIE 393
|||||
|||||

RESULT 12
ABU23355
ID ABU23355 standard; protein; 484 AA.
XX ABU23355;
XX 19-JUN-2003 (first entry)
XX
```

DE Protein encoded by Prokaryotic essential gene #8882.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Bordetella pertussis.

OS WO200277183-A2.

PN 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.

PF 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA27225.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 51279; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation; (7) identifying a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 484 AA;

Query Match 3.8%; Score 8; DB 6; Length 484;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 IEARGLKQ 83

Db 415 IEARGLKQ 422

RESULT 13

AAW68402

ID AAW68402 standard; protein; 580 AA.

XX AAW68402;

AC 28-OCT-1998 (first entry)

XX FeLV-A gag protein.

XX Multivalent vaccine; cat; pathogen; respiratory disease; FeLV; FPV; FCV;
KW digestive disease; feline leukemia virus; feline pan-leukemia virus;
KW feline calicivirus; feline immunodeficiency virus; FIV; rabies virus;
KW vector; envelope glycoprotein.

XX Feline leukemia virus.

XX FR2751223-A1.

XX 23-JAN-1998.

XX 19-JUL-1996; 96FR-00009337.

XX 19-JUL-1996; 96FR-00009337.

XX (INNR) RHONE MERIEUX SA.

XX Audonnet JCF, Bouchardon A, Baudu P, Riviere MEA;

XX WPI; 1998-112823/11.

XX N-PSDB; AAV49262.

XX Multi-valent polynucleotide vaccines against feline pathogens - consist
PT of at least 3 plasmids able to express protective antigens from specified
PT viruses.

XX Example 9; Fig 5; 42pp; French.

XX The invention relates to a multivalent vaccine for protecting cats
CC against several pathogens, especially pathogens associated with
CC respiratory and digestive diseases. The pathogens are especially selected
CC from feline leukemia virus (FeLV), feline pan-leukemia virus (FPV),
CC feline calicivirus (FCV), feline immunodeficiency virus (FIV), or rabies
CC virus. The vaccines are preferably composed of polynucleotide sequences
CC encoding 3 antigens, all as part of vectors. This sequence represents the
CC gag protein from the feline leukemia virus (FeLV) subtype A strain
CC Glasgow-1. The coding sequence was sub-cloned into the plasmid pVR1012 to
CC generate plasmid pPB181 for use in the vaccine

XX Sequence 580 AA;

Query Match 3.8%; Score 8; DB 2; Length 580;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 KEMTKVLA 203

Db 513 KEMTKVLA 520

RESULT 14

AAW50122

ID AAW50122 standard; protein; 611 AA.

XX AAW50122;

XX 21-DEC-2001 (first entry)

XX Feline leukemia virus p27-gp70 recombinant antigen pFeLVp27-gp70611.

XX FeLV; glycoprotein; gp70; p27; pFeLVp27-gp70611; antigen; immune status;

XX vaccination status; cat.
 XX Feline herpesvirus.
 OS Synthetic.
 XX WO20016568-A2.
 PN 13-SEP-2001.
 XX 07-MAR-2001; 2001WO-US007251.
 XX 09-MAR-2000; 2000US-00521738.
 PR (HESK-) HESKA CORP.
 PA (COLS) UNIV COLORADO STATE RES FOUND.
 XX Jensen WA, Lappin MR, Rosen DK, Andrews JS;
 XX WPI; 2001-639000/73.
 DR N-PSDB; AAH27069.
 XX Determining immune status or vaccination status of an animal to e.g.
 PT calcivirus comprises using a recombinant viral antigen.
 XX Claim 4; Page 119-121; 132pp; English.
 PS The present sequence is that of a fusion protein, termed pFelVp27-
 CC gp70611, between the C-terminus of feline leukemia virus (FeLV) pR65-gag
 CC and FeLV glycoprotein 70. The fusion protein was produced in Escherichia
 CC coli cells transformed by a recombinant vector comprising nucleic acid
 CC nFelVp27-gp701833 (see AAH27069). pFelVp27-gp70611 is an example of a
 CC recombinant infectious agent antigen that can be used in the method of
 CC the invention to determine the immune status of an animal. The method
 CC involves contacting a biological specimen of an animal (cat, dog or
 CC horse) with a recombinant antigen, and detecting the presence or absence
 CC of a complex between the recombinant antigen and an antibody present in
 CC the sample. The method determines whether the animal is protected against
 CC disease or should be vaccinated. Recombinant antigens (see AAH27069),
 CC nucleic acids encoding them (see AAH27054-71), methods of producing them,
 CC and assay methods are provided
 XX Sequence 611 AA;
 SQ Query Match 3.8%; Score 8; DB 4; Length 611;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 196 KEMTKVLA 203
 DB 131 KEMTKVLA 138
 |||||
 RESULT 15
 ABB92052
 ID ABB92052 standard; protein; 783 AA.
 XX ABB92052;
 AC 31-MAY-2002 (first entry)
 XX Herbicidally active polypeptide SEQ ID NO 1263.
 DT Herbicidal; plant; agriculture; herbicide.
 DE Arabidopsis thaliana.
 KW WO200210210-A2.
 OS 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP009892.
 PF 28-AUG-2001; 2001WO-EP009892.
 PR

XX (FARB) BAYER AG.
 PA Tietjen K, Weidler M;
 PI WPI; 2002-269010/31.
 DR Identifying plant target proteins for herbicidally active compounds,
 XX comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.
 PS Claim 5; SEQ ID NO 1263; 261pp + Sequence Listing; English.
 XX The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides
 XX Sequence 783 AA;
 SQ Query Match 3.8%; Score 8; DB 5; Length 783;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 KAIDDAIA 38
 DB 257 KAIDDAIA 264
 |||||
 RESULT 16
 AAR05898
 ID AAR05898 standard; protein; 1784 AA.
 XX AAR05898;
 AC 24-OCT-2003 (revised)
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 29-NOV-1990 (first entry)
 XX Gene product of first ORF of virus capable of inducing viraemia in feline
 DE species encoded by clone 61E.
 KW Viraemia; leukaemia; FeLV-A; clone 61C; clone 61E; ds.
 XX Feline leukemia virus; strain A.
 OS EP377842-A.
 PN 18-JUL-1990.
 PD 12-DEC-1989; 89EP-00122964.
 PF 13-DEC-1988; 88US-00284139.
 PR (HARD) UNIV HARVARD.
 PA (COLS) COLORADO STATE UNIV RES..
 PA (HARD) UNIV HARVARD.
 XX Hoover EA, Mullins JI;
 PI WPI; 1990-218326/29.
 DR N-PSDB; AAQ05252.
 XX DNA encoding leukaemia virus-A sub: type - producing AIDS type disease in
 PT cats, used to test drugs and vaccines.
 XX

PS Claim 6; Fig 1.1-1.7; 23pp; English.

XX Variant of viral genome induces immunodeficiency in cats similar to AIDS.
 CC May be used in research, especially in testing drugs and vaccines against
 CC viroemia and and feline leukaemia viruses. (Updated on 25-MAR-2003 to
 CC correct PA field.) (Updated on 27-AUG-2003 to correct OS field.) (Updated
 CC on 24-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1784 AA;

Query Match 3.8%; Score 8; DB 2; Length 1784;

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203

Db 513 KEMTKVLA 520

RESULT 17

AA94427
 ID AAR94427 standard; protein; 1784 AA.

XX AAR94427;

AC 16-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 11-JUN-1996 (first entry)

XX FELV FeA provirus clone 61E encoded nucleocapsid protein.
 DE
 XX
 XX FELV; retrovirus; vaccine; AIDS; disease model; immunodeficiency;
 KW viraemia; leukaemia; therapy; nucleocapsid.
 XX
 XX Feline immunodeficiency virus; subtype A.

XX Key Location/Qualifiers
 FH Misc-difference 581
 FT /note= "codon 581 in encoding sequence is TAG"
 FT
 XX
 XX EP699758-A1.
 PN
 XX
 XX PD 06-MAR-1996.
 XX
 XX PF 12-DEC-1989; 95EP-00100871.
 XX
 XX PR 13-DEC-1988; 88US-00284139.
 XX
 XX (HARD) UNIV HARVARD.
 PA (COLS) UNIV COLORADO STATE RES FOUND.
 XX
 XX PI Hoover EA, Mullins JI;
 XX
 XX WPI; 1996-180826/19.
 DR N-PSDB; AAT13265.
 CC
 XX
 XX Inactivated FeLV-A sub-type isolates for use in disease models and
 PT vaccines - can be used to study prophylaxis and therapy of related immuno
 PT -deficiencies in other species, e.g. human.
 XX
 XX Disclosure; Fig 1; 22pp; English.

XX The nucleocapsid protein (AAR94427) and envelope protein (AAR94428)
 CC sequences of feline leukaemia virus provirus FeA clone 61E were deduced
 CC from open reading frames identified in the proviral DNA (AAT13265). Clone
 CC 61E is replication competent and capable of inducing persistent viraemia
 CC in cats. Host cells, e.g. feline cells, can be transformed with DNA
 CC derived from 61E and used to produce infectious virus useful in vaccines,
 CC in the generation of viraemia and in disease challenge systems. (Updated
 CC on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to
 CC standardise OS field)
 CC
 XX Sequence 1784 AA;

Query Match

Best Local Similarity 3.8%; Score 8; DB 2; Length 1784;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 KEMTKVLA 203

Db 513 KEMTKVLA 520

RESULT 18

AA51943
 ID AAB51943 standard; protein; 73 AA.

XX AAB51943;

AC 21-FEB-2001 (first entry)

DT
 XX
 XX Human secreted protein sequence encoded by gene 15 SEQ ID NO:75.

DE
 XX
 XX Human; secreted protein; immunosuppressive; antiarthritic; vulnary;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; hyperproliferative disorder; neoplasm;
 KW autoimmune disease; rheumatoid arthritis; cardiovascular disorder;
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis; viral;
 KW nervous system disorder; Alzheimer's disease; bacteria; infection;
 KW cell proliferation; skin aging; wound healing; chemotaxis; food additive.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO2000058334-A1.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 22-MAR-2000; 2000WO-US007507.
 PF
 XX
 XX 26-MAR-1999; 99US-0126594P.
 PR
 XX 17-DEC-1999; 99US-0172408P.
 XX
 XX (HUYA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Rosen CA, Ruben SM, Komatsoulis G;
 PI
 XX
 XX WPI; 2000-611701/58.
 DR N-PSDB; AAC95476.
 DR
 XX
 XX New nucleic acid molecules encoding 50 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 PT
 XX
 XX Claim 11; Page 374; 402pp; English.

XX Polynucleotide sequences AAC95462 - AAC95511 represent cDNA encoding 50
 CC human secreted proteins AAB51929 - AAB51978. Sequences AAB51979 -
 CC AAB52010 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnary. The secreted
 CC proteins, polypeptides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The


```

XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS55546.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids.
XX PS
XX PS Example 3; SEQ ID NO 13280; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the genes,
XX CC their use in the discovery of novel antibiotics, the essential genes
XX CC themselves and the encoded proteins. The prokaryotes used are Escherichia
XX CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX CC useful for the identification of potential new targets for antibiotic
XX CC development. The antisense nucleic acids can also be used to identify
XX CC proteins used in proliferation, to express these proteins, and to obtain
XX CC antibodies capable of binding to the expressed proteins. The proteins can
XX CC be used to screen compounds in rational drug discovery programmes. The
XX CC antisense nucleic acid sequence is also useful to screen for homologous
XX CC of organisms. The present sequence represents an essential prokaryotic
XX CC cellular proliferation protein. Note: The sequence data for this patent
XX CC did not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 87 AA;

Query Match 3.3%; Score 7; DB 4; Length 87;
Best Local Similarity 100.0%; Pred.No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDAIAAI 40
Db 71 DDAIAAI 77

RESULT 21
ABP27648
ID ABP27648 standard; protein; 87 AA.
XX AC ABP27648;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 4472.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005540.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN68279.

```

```

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN56044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins
XX SQ Sequence 87 AA;

Query Match 3.3%; Score 7; DB 5; Length 87;
Best Local Similarity 100.0%; Pred.No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 DAIAAIE 41
Db 72 DAIAAIE 78

RESULT 22
ABU01601
ID ABU01601 standard; protein; 87 AA.
XX AC ABU01601;
XX DT 23-OCT-2003 (revised)
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #1177.
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
XX KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
XX KW gene therapy; vaccine.
XX OS Streptococcus pneumoniae; type 4 strain.
XX PN WO200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002WO-IB002163.
XX PR 27-MAR-2001; 2001GB-00007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Massignani V, Tettelin H, Fraser C;
XX DR WPI; 2003-040579/03.
XX DR N-PSDB; ABX06889.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX PT useful as medicaments for treating or preventing a disease or infection
XX PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX PT ear infection.

```

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA49935.
XX New antineoplastic nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 73989; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 87 AA;
Query Match 3.3%; Score 7; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 DDAIAAI 40
Db 71 DDAIAAI 77
RESULT 24
ABU44690
ID ABU44690 standard; protein; 87 AA.
XX ABU44690;
AC ABU44690;
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX Protein encoded by Prokaryotic essential gene #30217.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Streptococcus mutans.
OS
XX WO200277183-A2.
XX
XX 03-OCT-2002.
PD
XX

XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX ABS56454. Also included are an antibody which binds one of the proteins,
XX treating a patient by administering the protein, DNA or antibody (in a
XX composition), a kit comprising first and second primers, which are the
XX nucleic acid cited above or fragments between nucleotides 8-100 of a
XX sequence not defined in the specification, for amplifying a target
XX sequence contained within a Streptococcus nucleic acid sequence, where
XX the first primer is substantially complementary to the target sequence
XX and the second primer is substantially complementary to the complement of
XX the target sequence, and where the parts of the primers having
XX substantial complementarity define the termini of the target sequence to
XX be amplified, assay comprising contacting a test compound with the
XX protein, and determining whether the test compound binds to the protein
XX and a Streptococcus pneumoniae bacterium, where one or more genes
XX encoding the proteins has been rendered inactive. The proteins, nucleic
XX acid molecules, antibody and compositions are useful as medicaments for
XX treating or preventing a disease or infection due to streptococcus
XX bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis
XX media or ear infection. They are also useful in developing vaccines,
XX diagnostics and antibiotics. The methods are useful for identifying
XX immunodominant proteins. The present sequence is one of the 2469 proteins
XX expressed by the identified coding regions from the genomic sequence.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
XX standardise OS field)
XX
XX Sequence 87 AA;
Query Match 3.3%; Score 7; DB 6; Length 87;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 DDAIAAI 40
Db 71 DDAIAAI 77
RESULT 23
ABU46065
ID ABU46065 standard; protein; 87 AA.
XX ABU46065;
AC ABU46065;
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX Protein encoded by Prokaryotic essential gene #31592.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Streptococcus pneumoniae.
OS
XX WO200277183-A2.
XX
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00815242.
PR
XX 06-SEP-2001; 2001US-00948993.
PR
XX 25-OCT-2001; 2001US-0342923P.
PR
XX 08-FEB-2002; 2002US-00072851.
PR
XX 06-MAR-2002; 2002US-0362693P.
PR
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX

PF XX 21-MAR-2002; 2002WO-US009107.
 PR XX 21-MAR-2001; 2001US-00815242.
 PR XX 06-SEP-2001; 2001US-00948993.
 PR XX 25-OCT-2001; 2001US-0342923P.
 PR XX 08-FEB-2002; 2002US-00072851.
 PR XX 06-MAR-2002; 2002US-0362699P.
 XX XX (ELIT-) ELITRA PHARM INC.
 XX XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 DR XX WPI; 2003-029926/02.
 DR XX N-PSDB; ACA48560.
 XX XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids, required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX XX Claim 25; SEQ ID NO 72614; 1765pp; English.
 XX XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX XX Sequence 87 AA;
 SQ
 Query Match 3.3%; Score 7; DB 6; Length 87;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 DDAIAAI 40
 Db 71 DDAIAAI 77
 RESULT 25
 ABU46703
 ID ABU46703 standard; protein; 87 AA.
 AC ABU46703;
 XX
 DT 19-JUN-2003 (first entry)
 XX

DE XX Protein encoded by Prokaryotic essential gene #32230.
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Streptococcus pyogenes.
 OS
 XX WO200277183-A2.
 PN
 XX 03-OCT-2002.
 PD
 XX 21-MAR-2002; 2002WO-US009107.
 PF
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX XX (ELIT-) ELITRA PHARM INC.
 XX XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 PI XX WPI; 2003-029926/02.
 DR XX N-PSDB; ACA50573.
 DR XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX XX Claim 25; SEQ ID NO 74627; 1766pp; English.
 PS The invention relates to an isolated nucleic acid comprising any one of
 XX the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX XX Sequence 87 AA;
 SQ
 Query Match 3.3%; Score 7; DB 6; Length 87;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 35 DAAIAIE 41
 Db 72 DAAIAIE 78

RESULT 26

AAG01133
ID AAG01133 standard; protein; 133 AA.

XX AC AAG01133;
XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein, SEQ ID NO: 5214.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.

XX OS Homo sapiens.
XX FN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-00200610.
XX PR 26-FEB-1999; 99US-0122487P.

XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR N-PSDB; AAC01139.

XX KW New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX KW obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX KW diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX PS Claim 13; SEQ ID NO 5214; 71pp + Sequence Listing; English.

XX CC The present sequence is a polypeptide encoded by one of a large number of
XX CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
XX CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
XX CC tissues. EST sequences usually correspond mainly to the 3' untranslated
XX CC region (UTR) of the mRNA because they are often obtained from oligo-dT
XX CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
XX CC sequences derived from the 5' ends of mRNAs and even in those cases where
XX CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
XX CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
XX CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
XX CC are also used in diagnostic, forensic, gene therapy and chromosome
XX CC mapping procedures. They are used to obtain upstream regulatory sequences
XX CC and to design expression and secretion vectors

XX SQ Sequence 133 AA;

Query Match 3.3%; Score 7; DB 3; Length 133;
Best Local Similarity 100.0%; Pred.No.1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 27

AAU30907
ID AAU30907 standard; protein; 139 AA.

XX AC AAU30907;
XX DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #1398.

XX KW Human; vaccination; gene therapy; nutritional supplement;
XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.
XX FN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US008656.
XX PR 18-APR-2000; 2000US-00552929.
XX PR 26-JAN-2001; 2001US-00770160.
XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-611725/70.
XX KW Nucleic acids encoding a range of human polypeptides, useful in genetic
XX KW vaccination, testing and therapy.

XX PS Claim 20; Page 368; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising the
XX CC nucleic acids encoding the polypeptides and cells genetically engineered
XX CC to express them are also useful for producing the proteins. The proteins
XX CC are useful in genetic vaccination, testing and therapy, and can be used
XX CC as nutritional supplements. They may be used to increase stem cell
XX CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
XX CC and/or nerve tissue growth or regeneration; immune suppression and/or
XX CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.

XX CC AAU29510-AAU3304 represent the amino acid sequences of novel human
XX CC secreted proteins of the invention

XX SQ Sequence 139 AA;

Query Match 3.3%; Score 7; DB 4; Length 139;
Best Local Similarity 100.0%; Pred.No.1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19
DB 59 VAVSADP 65

RESULT 28

ADA33025
ID ADA33025 standard; protein; 148 AA.

XX AC ADA33025;
XX DT 20-NOV-2003 (first entry)

XX DE Acinetobacter baumannii protein #186.
XX KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX KW plant biocontrol agent.

XX OS Acinetobacter baumannii.
XX PN US6562958-B1.
XX XX 13-MAY-2003.

XX PD 13-MAY-2003.

```
XX 04-JUN-1999; 99US-00328352.
PF 09-JUN-1998; 98US-0088701.P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Breton G, Bush D;
PI WPI; 2003-576092/54.
XX N-PSDB; ADA28899.
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX Example; SEQ ID NO 4312; 328pp; English.
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents.
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX Sequence 148 AA;
SQ
Query Match 3.3%; Score 7; DB 6; Length 148;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 169 IGGLSIL 175
DB 21 IGGLSIL 27
RESULT 29
AAU30909
ID AAU30909 standard; protein; 152 AA.
XX AAU30909;
AC
XX 18-DEC-2001 (first entry)
DT
XX Novel human secreted protein #1400.
DE
XX Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX Homo sapiens.
OS
XX WO200179449-A2.
FN
XX 25-OCT-2001.
PD
XX 16-APR-2001; 2001WO-US0088566.
PF
XX 18-APR-2000; 2000US-00552929.
PR
XX 26-JAN-2001; 2001US-00770160.
PR (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-611725/70.
DR
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
```

```
XX Claim 20; Page 369-370; 765pp; English.
XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. the polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX Sequence 152 AA;
SQ
Query Match 3.3%; Score 7; DB 4; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 VAVSADP 19
DB 27 VAVSADP 33
RESULT 30
AAG61764
ID AAG61764 standard; protein; 155 AA.
XX AAG61764;
AC
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 80209.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
FN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-00301439.
PF
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
```

Mon Aug 9 08:14:01 2004

us-10-024-955-7.oligo.rag

PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136399P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140895P.
PR 28-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 08-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147202P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154038P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.

PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161408P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.
 PR 29-OCT-1999; 99US-0162143P.

Query Match 3.3%; Score 7; DB 3; Length 155;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAIFQ 191
 |||||
 DB 115 VLTAIFQ 121
 |||||

RESULT 31
 ABP26815
 ID ABP26815 standard; protein; 178 AA.
 AC ABP26815;
 XX
 XX
 DT 02-JUL-2002 (first entry)
 DE Streptococcus polypeptide SEQ ID NO 2806.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; Infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 XX WO200234771-A2.
 XX
 XX
 PD 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB004789.
 XX
 XX 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 XX WPI; 2002-352536/38.
 DR N-PSDB; ABN67445.
 XX
 XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX
 PS Claim 1; Page 3431; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (i), nucleic acids encoding (i), ABN66044-ABN71526 and
 CC antibodies that bind (i) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (i) are used to detect Streptococcus in a
 CC biological sample. (i) is used to determine whether a compound binds to
 CC (i). A composition comprising (i) or a nucleic acid encoding (i), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (i) may be used to recombinantly produce (i) and may be
 CC used in gene therapy. Antibodies to (i) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX
 XX Sequence 178 AA;
 QY 179 FGVLSDV 185
 |||||
 DB 123 FGVLSDV 129
 |||||

RESULT 32
 ADD71115
 ID ADD71115 standard; protein; 178 AA.
 XX
 AC ADD71115;
 XX
 DT 15-JAN-2004 (first entry)
 DE Human intracellular signalling molecule INTSIG-4 protein SEQ ID NO:4.
 XX
 KW human; intracellular signalling molecule; INTSIG; cytosolic;
 KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;
 KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
 KW antiinflammatory; thymomimetic; gene therapy;
 KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;
 KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;
 KW immune disorder; inflammatory disorder; AIDS; allergy;
 KW developmental disorder; Hypothyroidism; Cushing's syndrome;
 KW reproductive disorder; vesicle-trafficking disorder; infection.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO2003039348-A2.
 XX
 XX 15-MAY-2003.
 XX
 XX 07-NOV-2002; 2002WO-US036151.
 XX
 XX 09-NOV-2001; 2001US-0344472P.
 PR 30-NOV-2001; 2001US-0334558P.
 PR 14-DEC-2001; 2001US-0340296P.
 PR 21-DEC-2001; 2001US-0343557P.
 PR 18-JAN-2002; 2002US-0350420P.
 PR 25-JAN-2002; 2002US-0351927P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Kable AB, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;
 PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;
 PI Marquis JP, Gietzen KU, Baughn MR, Borowsky ML, Yao MG, Chawla NK;
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Satra S, Becha SD;
 PI Lee SY, Tran UK, Elliott VS, Sprague WW, Tang Y, Zebardjian Y;
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;
 PI Zheng W;
 XX
 XX WPI; 2003-441441/41.
 DR N-PSDB; ADD71167.
 XX
 XX New human intracellular signalling molecules (INTSIG), useful for
 PT diagnosing, treating and preventing diseases or conditions associated

PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or
PT infections.
XX Claim 1; SEQ ID NO 4; 363pp; English.
PS
XX The present invention describes human intracellular signalling molecules
CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins
CC have cytosolic, antiarteriosclerotic, antidiabetic, anticonvulsant,
CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
CC antiinflammatory and thymomimetic activities, and can be used in gene
CC therapy. The INTSIG polynucleotides and proteins can be used in the
CC diagnosis, treatment and prevention of diseases or conditions associated
CC with the decreased expression or overexpression of INTSIG, such as cell
CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),
CC neurological (e.g. epilepsy, Huntington's disease, stroke),
CC immune/inflammatory (e.g. AIDS, allergies), developmental (e.g.
CC hypochondrodysplasia, Cushing's syndrome), reproductive and vesicle-trafficking
CC disorders, or infections. They are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of INTSIG. The INTSIG or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The present sequence represents human INTSIG-4 from the present
CC invention.
XX
XX
SQ Sequence 178 AA;
Query Match 3.3%; Score 7; DB 7; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
QY 13 VAYSADP 19
Db 59 VAYSADP 65
|||||
RESULT 33
AAG59755
ID AAG59755 standard; protein; 181 AA.
XX
AC AAG59755;
DT
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 77324.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136031P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137744P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.

```
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154015P.
PR 16-SEP-1999; 99US-0154035P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155135P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155653P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154015P.
PR 16-SEP-1999; 99US-0154035P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155135P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155653P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
Query Match 3.3%; Score 7; DB 3; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 VLTAIFQ 191
Db 83 VLTAIFQ 89
RESULT 34
AAG59754
ID AAG59754 standard; protein; 183 AA.
XX
AC AAG59754;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 77323.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
```

PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132456P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144844P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159239P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.


```
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147360P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149802P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 185 VLTAIFQ 191
Db 87 VLTAIFQ 93
|||||

RESULT 36
AAG55324
ID AAG55324 standard; protein; 185 AA.
XX
AC AAG55324;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 70920.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127452P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
```

PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0148368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 07-SEP-1999; 99US-0153070P.
PR 10-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 13-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 185 VLTAIFQ 191
DB 88 VLTAIFQ 94
|||||

RESULT 37
ABP69706
ID ABP69706 standard; protein; 190 AA.
XX
AC ABP69706;
XX
XX 20-JAN-2003 (first entry)
XX
XX Human polypeptide SEQ ID NO 1753.
XX
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic.
XX
OS Homo sapiens.
XX
XX WO200270539-A2.
XX
XX 12-SEP-2002.
XX
XX 05-MAR-2002; 2002WO-US005095.
XX
XX 05-MAR-2001; 2001US-00799451.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX
XX WPI; 2002-759812/82.
XX
XX N-PSDB; ABZ11923.
XX
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.
XX
XX Claim 9; SEQ ID NO 1753; 1012pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences (ABZ1119-
CC ABZ12066) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 190 AA;

Query Match 3.3%; Score 7; DB 5; Length 190;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 VAVSADP 19
DB 59 VAVSADP 65
|||||

RESULT 38
ADA24313
ID ADA24313 standard; protein; 192 AA.
XX
AC ADA24313;
XX
XX 20-NOV-2003 (first entry)
XX
XX Bacillus clausii SipV protein partial sequence.
XX
XX secretion factor; SF; SecA; SecD; SecE; SecF; SecG; SecY; Ffh; FtsY;
KW SipS; SipT; SipV; SipW; transcription gene; sporulation gene;
KW sporulation deficient; Gram positive microorganism; Bacillus genus;
KW large-scale industrial fermentation; fermentation product;
KW sporulation factor; industrial protein; therapeutic protein;
KW agricultural protein; food; animal feed; cosmetic; cleaning composition;
KW enzyme; hormone; growth factor.
XX
XX Bacillus clausii.
OS
XX
XX WO2003066818-A2.
XX
XX 14-AUG-2003.
XX
XX 06-FEB-2003; 2003WO-US003534.
XX
XX 08-FEB-2002; 2002US-0355259P.
XX
XX (GEMV) GENENCOR INT INC.
XX
XX Ferrari E, Van Kimmenade A;
XX
XX WPI; 2003-646305/61.
XX
XX N-PSDB; ADA24314.
XX
XX New nucleic acid from Bacillus clausii, useful for improving recombinant
PT protein production, encodes secretion, transcription or sporulation
PT factors, also encoded proteins.
XX
XX Claim 11; Fig 29; 67pp; English.
XX
XX This invention relates to novel DNA sequences which encode one of the
CC Bacillus clausii secretion factors (SF) SecA, SecD, SecE, SecF, SecG,
CC SecY, Ffh, FtsY, SipS, SipT, SipV and SipW. In addition, the invention
CC relates to novel transcription and sporulation genes in Bacillus clausii.
CC It is desirable that a production strain of Bacillus clausii should be
CC deficient in sporulation, hence a greater understanding of this process
CC is proposed by the identification of the genes of the invention. Gram
CC positive microorganisms, such as members of the genus Bacillus, have been
CC used for large-scale industrial fermentation due in part, to their
CC ability to secrete their fermentation products into their culture media.
CC Increasing the ability of these microorganisms to express and secrete
CC polypeptides is therefore of great interest in many fields. The secretion
CC of polypeptides by members of the genus Bacillus may be further
CC understood and increased by making use of the proteins and DNA sequences
CC of the invention. Sequences from Bacillus clausii that encode
CC transcription or sporulation factors are used in recombinant expression
CC and secretion of proteins, for example those for industrial, therapeutic
CC or agricultural use, also in food, animal feed, cosmetics and cleaning
CC compositions, typically enzymes, hormones and growth factors.
CC Transformation with sequences that encode secretion, transcription or
CC sporulation factors may increase both production and secretion of
CC proteins in Bacillus. The present sequence is the partial amino acid
CC sequence of the Bacillus clausii SipV protein of the invention.

XX SQ Sequence 192 AA;
 Query Match 3.3%; Score 7; DB 7; Length 192;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 MTSFEVR 160
 |||||
 DB 44 MTSFEVR 50

RESULT 39
 ABB06132
 ID ABB06132 standard; protein; 193 AA.
 XX AC ABB06132;
 DT 10-MAY-2002 (first entry)
 XX DE Human NS protein sequence SEQ ID NO:224.
 XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;
 KW anti-rheumatic; antiarthritic; antipsoriatic; ophthalmological; anti-HIV;
 KW vasoregic; antiarteriosclerotic; antiinflammatory; dermatological;
 KW anorectic; muscular; antiinfertility; cardiovascular; anticoagulant;
 KW antifibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiant;
 KW anticonvulsant; antidiabetic; tranquiliser; antidepressant; neurotropic;
 KW gastrointestinal; virucide; antiulcer; cancer; osteoporosis; dystonia;
 KW contraceptive; vaccine; gene therapy; multiple sclerosis; psoriasis;
 KW endometriosis; degenerative disease; atherosclerosis; glaucoma;
 KW rheumatoid arthritis; cataract; restenosis; muscular dystrophy; AIDS;
 KW inflammation; skin disorder; obesity; conglutination disease; hypertension;
 KW ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;
 KW diabetes; anxiety; depression; schizophrenia; viral disease; stroke;
 KW gastric ulcer; Alzheimer's disease.
 OS Homo sapiens.
 PN WO200206315-A2.
 XX PD 24-JAN-2002.
 XX PF 17-JUL-2001; 2001WO-IL000653.
 XX PR 18-JUL-2000; 2000IL-00137345.
 XX PR 15-DEC-2000; 2000IL-00140354.
 XX PA (COMP-) COMPUGEN LTD.
 XX PI Mintz L, Freilich S, Bernstein J;
 XX WPI: 2002-155037/20.
 XX DR N-PSDB; ABI39786.
 XX One hundred and twenty eight novel nucleic acid sequences, useful for
 treating and diagnosing e.g. cancer, asthma and Alzheimer's.
 PS Claim 6; Page 260-261; 290pp; English.
 CC ABL39691 to ABL39918 represent novel human nucleic acid sequences
 encoding the proteins given in ABB06037 to ABB06164. The novel sequences
 (NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,
 CC anti-rheumatic, antiarthritic, antipsoriatic, ophthalmological, virucide,
 CC vasoregic, antiarteriosclerotic, antiinflammatory, dermatological,
 CC anorectic, muscular, anti-HIV, antiinfertility, cardiovascular,
 CC anticoagulant, antifibrinolytic, hypotension, antiasthmatic, cardiant,
 CC immunomodulator, anticonvulsant, antidiabetic, tranquiliser, antiulcer,
 CC antidepressant, gastrointestinal, antiepileptic, cerebroprotective,
 CC neurotropic and contraceptive activities. The NS can be used in vaccines,
 CC gene therapy and antisense therapy. Nucleic acids, expression vectors and
 CC antibodies from the present invention can be used for treating and

CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative
 CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
 CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
 CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
 CC disease, coagulation disease, ischaemia, hypertension, asthma, immune
 CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
 CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
 CC Alzheimer's disease and as a contraceptive
 XX SQ Sequence 193 AA;
 Query Match 3.3%; Score 7; DB 5; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
 |||||
 DB 59 VAVSADP 65

RESULT 40
 AAR54114
 ID AAR54114 standard; protein; 204 AA.
 XX AC AAR54114;
 DT 25-MAR-2003 (revised)
 DT 20-JAN-1995 (first entry)
 XX DE GDI D4 protein.
 XX D4; GDP-dissociation inhibitor; GDI; hematopoietic cell; diagnostic;
 KW leukemia; tumor bone marrow transplant.
 OS Bos taurus.
 XX WO9413802-A1.
 XX PD 23-JUN-1994.
 XX PF 10-DEC-1993; 93WO-US012074.
 XX PR 10-DEC-1992; 92US-00990337.
 XX (CYTO-) CYTOMED INC.
 XX (BETH-) BETH ISRAEL HOSPITAL ASSOC.
 XX Lim B, Lelias J, Adra CN, Ko JL;
 XX WPI: 1994-217878/26.
 XX New D4 protein preferentially expressed in hematopoietic cells - for use
 as a diagnostic, in research and for treatment of leukaemia tumours, and
 in bone marrow transplants.
 PS Disclosure; Page 56-57; 76pp; English.
 XX Differential screening of a subtractive hematopoietic cDNA library
 yielded D4 cDNA, which encoded a protein that showed GDP-dissociation
 CC inhibitor activity. The cDNA sequence was determined (AAQ68715) and the
 CC amino acid sequence corresponding to the longest ORF was deduced
 CC (AAR54112). Equivalent sequences for mouse D4 are given in AAQ68716 and
 CC (AAR54113). The cattle D4 protein sequence was also provided (AAR54114).
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 204 AA;
 Query Match 3.3%; Score 7; DB 2; Length 204;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
 |||||

Db	59	VAVSADP 65	PR	18-JUN-1999;	99US-0139456P.
			PR	18-JUN-1999;	99US-0139457P.
			PR	18-JUN-1999;	99US-0139458P.
			PR	18-JUN-1999;	99US-0139459P.
			PR	18-JUN-1999;	99US-0139460P.
			PR	18-JUN-1999;	99US-0139461P.
			PR	18-JUN-1999;	99US-0139462P.
			PR	18-JUN-1999;	99US-0139463P.
			PR	18-JUN-1999;	99US-0139750P.
			PR	18-JUN-1999;	99US-0139763P.
			PR	21-JUN-1999;	99US-0139817P.
			PR	22-JUN-1999;	99US-0139899P.
			PR	23-JUN-1999;	99US-0140353P.
			PR	23-JUN-1999;	99US-0140354P.
			PR	24-JUN-1999;	99US-0140698P.
			PR	28-JUN-1999;	99US-0140823P.
			PR	29-JUN-1999;	99US-0140911P.
			PR	30-JUN-1999;	99US-0141287P.
			PR	01-JUL-1999;	99US-0141842P.
			PR	01-JUL-1999;	99US-0142154P.
			PR	02-JUL-1999;	99US-0142055P.
			PR	06-JUL-1999;	99US-0142390P.
			PR	08-JUL-1999;	99US-0142803P.
			PR	09-JUL-1999;	99US-0142920P.
			PR	12-JUL-1999;	99US-0142977P.
			PR	13-JUL-1999;	99US-0143542P.
			PR	14-JUL-1999;	99US-0143624P.
			PR	15-JUL-1999;	99US-0144005P.
			PR	16-JUL-1999;	99US-0144085P.
			PR	16-JUL-1999;	99US-0144086P.
			PR	19-JUL-1999;	99US-0144331P.
			PR	19-JUL-1999;	99US-0144332P.
			PR	19-JUL-1999;	99US-0144333P.
			PR	19-JUL-1999;	99US-0144334P.
			PR	19-JUL-1999;	99US-0144335P.
			PR	20-JUL-1999;	99US-0144352P.
			PR	20-JUL-1999;	99US-0144632P.
			PR	20-JUL-1999;	99US-0144884P.
			PR	21-JUL-1999;	99US-0144814P.
			PR	21-JUL-1999;	99US-0145086P.
			PR	22-JUL-1999;	99US-0145088P.
			PR	22-JUL-1999;	99US-0145089P.
			PR	22-JUL-1999;	99US-0145087P.
			PR	23-JUL-1999;	99US-0145192P.
			PR	23-JUL-1999;	99US-0145145P.
			PR	23-JUL-1999;	99US-0145218P.
			PR	26-JUL-1999;	99US-0145224P.
			PR	26-JUL-1999;	99US-0145216P.
			PR	27-JUL-1999;	99US-0145913P.
			PR	27-JUL-1999;	99US-0145918P.
			PR	27-JUL-1999;	99US-0145919P.
			PR	28-JUL-1999;	99US-0145951P.
			PR	02-AUG-1999;	99US-0146386P.
			PR	02-AUG-1999;	99US-0146388P.
			PR	02-AUG-1999;	99US-0146389P.
			PR	03-AUG-1999;	99US-0147038P.
			PR	04-AUG-1999;	99US-0147204P.
			PR	04-AUG-1999;	99US-0147302P.
			PR	05-AUG-1999;	99US-0147192P.
			PR	05-AUG-1999;	99US-0147260P.
			PR	06-AUG-1999;	99US-0147303P.
			PR	06-AUG-1999;	99US-0147416P.
			PR	09-AUG-1999;	99US-0147493P.
			PR	09-AUG-1999;	99US-0147935P.
			PR	10-AUG-1999;	99US-0148171P.
			PR	11-AUG-1999;	99US-0148319P.
			PR	13-AUG-1999;	99US-0148341P.
			PR	13-AUG-1999;	99US-0148565P.
			PR	13-AUG-1999;	99US-0148684P.
			PR	16-AUG-1999;	99US-0149368P.
			PR	17-AUG-1999;	99US-0149175P.

RESULT 41
AAG59753
ID AAG59753 standard; protein; 205 AA.
XX AC
XX AAG59753;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 77322.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 77322.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-01322407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 18-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.

PR	18-AUG-1999;	99US-0149426P.	AC	AAG55323;
PR	20-AUG-1999;	99US-0149722P.	XX	18-OCT-2000 (first entry)
PR	20-AUG-1999;	99US-0149723P.	DT	
PR	20-AUG-1999;	99US-0149929P.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 70919.
PR	23-AUG-1999;	99US-0149902P.	DE	
PR	23-AUG-1999;	99US-0149930P.	XX	Protein identification; signal transduction pathway; metabolic pathway;
PR	25-AUG-1999;	99US-0150566P.	KW	hybridisation assay; Genetic mapping; gene expression control; promoter;
PR	26-AUG-1999;	99US-0150884P.	KW	termination sequence.
PR	27-AUG-1999;	99US-0151065P.	XX	
PR	27-AUG-1999;	99US-0151066P.	OS	Arabidopsis thaliana.
PR	27-AUG-1999;	99US-0151080P.	XX	EP1033405-A2.
PR	31-AUG-1999;	99US-0151303P.	PN	
PR	31-AUG-1999;	99US-0151438P.	XX	06-SEP-2000.
PR	01-SEP-1999;	99US-0151930P.	XX	
PR	07-SEP-1999;	99US-0152363P.	XX	
PR	10-SEP-1999;	99US-0153070P.	PF	25-FEB-2000; 2000EP-00301439.
PR	13-SEP-1999;	99US-0153758P.	XX	25-FEB-1999; 99US-0121825P.
PR	15-SEP-1999;	99US-0154018P.	PR	05-MAR-1999; 99US-0123180P.
PR	16-SEP-1999;	99US-0154039P.	PR	09-MAR-1999; 99US-0123548P.
PR	20-SEP-1999;	99US-0154779P.	PR	23-MAR-1999; 99US-0125788P.
PR	22-SEP-1999;	99US-0155139P.	PR	25-MAR-1999; 99US-0126264P.
PR	23-SEP-1999;	99US-0155486P.	PR	29-MAR-1999; 99US-0126785P.
PR	24-SEP-1999;	99US-0155653P.	PR	01-APR-1999; 99US-0127462P.
PR	28-SEP-1999;	99US-0156458P.	PR	06-APR-1999; 99US-0128234P.
PR	29-SEP-1999;	99US-0156596P.	PR	08-APR-1999; 99US-0128714P.
PR	04-OCT-1999;	99US-0157117P.	PR	16-APR-1999; 99US-0129845P.
PR	06-OCT-1999;	99US-0157865P.	PR	19-APR-1999; 99US-0130077P.
PR	07-OCT-1999;	99US-0158029P.	PR	21-APR-1999; 99US-0130449P.
PR	08-OCT-1999;	99US-0158234P.	PR	23-APR-1999; 99US-0130510P.
PR	12-OCT-1999;	99US-0158369P.	PR	22-APR-1999; 99US-0130891P.
PR	13-OCT-1999;	99US-0159293P.	PR	28-APR-1999; 99US-0131443P.
PR	13-OCT-1999;	99US-0159294P.	PR	30-APR-1999; 99US-0132048P.
PR	14-OCT-1999;	99US-0159329P.	PR	30-APR-1999; 99US-0132407P.
PR	14-OCT-1999;	99US-0159330P.	PR	04-MAY-1999; 99US-0132484P.
PR	14-OCT-1999;	99US-0159331P.	PR	05-MAY-1999; 99US-0132485P.
PR	14-OCT-1999;	99US-0159637P.	PR	06-MAY-1999; 99US-0132486P.
PR	14-OCT-1999;	99US-0159638P.	PR	07-MAY-1999; 99US-0132487P.
PR	18-OCT-1999;	99US-0159584P.	PR	07-MAY-1999; 99US-0132863P.
PR	21-OCT-1999;	99US-0160741P.	PR	11-MAY-1999; 99US-0134256P.
PR	21-OCT-1999;	99US-0160767P.	PR	14-MAY-1999; 99US-0134218P.
PR	21-OCT-1999;	99US-0160768P.	PR	14-MAY-1999; 99US-0134219P.
PR	21-OCT-1999;	99US-0160770P.	PR	14-MAY-1999; 99US-0134221P.
PR	21-OCT-1999;	99US-0160814P.	PR	14-MAY-1999; 99US-0134370P.
PR	21-OCT-1999;	99US-0160815P.	PR	18-MAY-1999; 99US-0134768P.
PR	22-OCT-1999;	99US-0160980P.	PR	19-MAY-1999; 99US-0134941P.
PR	22-OCT-1999;	99US-0160981P.	PR	20-MAY-1999; 99US-0135124P.
PR	22-OCT-1999;	99US-0160989P.	PR	21-MAY-1999; 99US-0135353P.
PR	25-OCT-1999;	99US-0161404P.	PR	24-MAY-1999; 99US-0135629P.
PR	25-OCT-1999;	99US-0161405P.	PR	25-MAY-1999; 99US-0136021P.
PR	25-OCT-1999;	99US-0161406P.	PR	27-MAY-1999; 99US-0136392P.
PR	26-OCT-1999;	99US-0161359P.	PR	28-MAY-1999; 99US-0136782P.
PR	26-OCT-1999;	99US-0161360P.	PR	01-JUN-1999; 99US-0137222P.
PR	26-OCT-1999;	99US-0161361P.	PR	03-JUN-1999; 99US-0137528P.
PR	28-OCT-1999;	99US-0161920P.	PR	04-JUN-1999; 99US-0137502P.
PR	28-OCT-1999;	99US-0161922P.	PR	07-JUN-1999; 99US-0137724P.
PR	28-OCT-1999;	99US-0161993P.	PR	08-JUN-1999; 99US-0138094P.
PR	28-OCT-1999;	99US-0161994P.	PR	10-JUN-1999; 99US-0138540P.
PR	29-OCT-1999;	99US-0162142P.	PR	10-JUN-1999; 99US-0138847P.
PR	29-OCT-1999;	99US-0162143P.	PR	14-JUN-1999; 99US-0139119P.
PR	29-OCT-1999;	99US-0162144P.	PR	16-JUN-1999; 99US-0139452P.
PR	29-OCT-1999;	99US-0162145P.	PR	16-JUN-1999; 99US-0139453P.
PR	29-OCT-1999;	99US-0162146P.	PR	17-JUN-1999; 99US-0139454P.
PR	29-OCT-1999;	99US-0162147P.	PR	18-JUN-1999; 99US-0139455P.
PR	29-OCT-1999;	99US-0162148P.	PR	18-JUN-1999; 99US-0139456P.
PR	29-OCT-1999;	99US-0162149P.	PR	18-JUN-1999; 99US-0139457P.
PR	29-OCT-1999;	99US-0162150P.	PR	18-JUN-1999; 99US-0139458P.
PR	29-OCT-1999;	99US-0162151P.	PR	18-JUN-1999; 99US-0139459P.
PR	29-OCT-1999;	99US-0162152P.	PR	18-JUN-1999; 99US-0139460P.
PR	29-OCT-1999;	99US-0162153P.	PR	18-JUN-1999; 99US-0139461P.
PR	29-OCT-1999;	99US-0162154P.	PR	18-JUN-1999; 99US-0139462P.

Query Match 3.3%; Score 7; DB 3; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAFQ 191
Db 107 VLTAFQ 113

PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.

PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154773P.
PR 22-SEP-1999; 99US-0155133P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAIFQ 191
|||
Db 117 VLTAIFQ 123

RESULT 43
AAG55906
ID AAG55906 standard; protein; 215 AA.
XX
AC AAG55906;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 71760.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

OS Arabidopsis thaliana.

XX EP103405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PF 25-FEB-1999; 99US-0121825P.

XX PR 09-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0129845P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 30-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 04-MAY-1999; 99US-0132407P.

XX PR 04-MAY-1999; 99US-0132484P.

XX PR 05-MAY-1999; 99US-0132485P.

XX PR 06-MAY-1999; 99US-0132486P.

XX PR 07-MAY-1999; 99US-0132487P.

XX PR 07-MAY-1999; 99US-0132862P.

XX PR 11-MAY-1999; 99US-0134256P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134219P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 18-MAY-1999; 99US-0134768P.

XX PR 19-MAY-1999; 99US-0134941P.

XX PR 20-MAY-1999; 99US-0135124P.

XX PR 21-MAY-1999; 99US-0135353P.

XX PR 24-MAY-1999; 99US-0135629P.

XX PR 25-MAY-1999; 99US-0136021P.

XX PR 27-MAY-1999; 99US-0136392P.

XX PR 28-MAY-1999; 99US-0136782P.

XX PR 01-JUN-1999; 99US-0137222P.

XX PR 03-JUN-1999; 99US-0137528P.

XX PR 04-JUN-1999; 99US-0137502P.

XX PR 07-JUN-1999; 99US-0137724P.

XX PR 08-JUN-1999; 99US-0138094P.

XX PR 10-JUN-1999; 99US-0138540P.

XX PR 10-JUN-1999; 99US-0138847P.

XX PR 14-JUN-1999; 99US-0139119P.

XX PR 16-JUN-1999; 99US-0139452P.

XX PR 16-JUN-1999; 99US-0139453P.

XX PR 17-JUN-1999; 99US-0139492P.

XX PR 18-JUN-1999; 99US-0139455P.

XX PR 18-JUN-1999; 99US-0139456P.

XX PR 18-JUN-1999; 99US-0139457P.

XX PR 18-JUN-1999; 99US-0139458P.

XX PR 18-JUN-1999; 99US-0139459P.

XX PR 18-JUN-1999; 99US-0139460P.

XX PR 18-JUN-1999; 99US-0139461P.

XX PR 18-JUN-1999; 99US-0139462P.

XX PR 18-JUN-1999; 99US-0139463P.

XX PR 18-JUN-1999; 99US-0139750P.

XX PR 18-JUN-1999; 99US-0139763P.

XX PR 21-JUN-1999; 99US-0139817P.

XX PR 22-JUN-1999; 99US-0139899P.

XX PR 23-JUN-1999; 99US-0140353P.

XX PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.

PR 21-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.

PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.

PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.

PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.

PR 02-AUG-1999; 99US-0146389P.

PR 03-AUG-1999; 99US-0147038P.

PR 04-AUG-1999; 99US-0147204P.

PR 04-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147192P.

PR 05-AUG-1999; 99US-0147260P.

PR 06-AUG-1999; 99US-0147303P.

PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.

PR 10-AUG-1999; 99US-0148171P.

PR 11-AUG-1999; 99US-0148319P.

PR 12-AUG-1999; 99US-0148341P.

PR 13-AUG-1999; 99US-0148565P.

PR 13-AUG-1999; 99US-0148684P.

PR 16-AUG-1999; 99US-0149368P.

PR 17-AUG-1999; 99US-0149175P.

PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.

PR 20-AUG-1999; 99US-0149723P.

PR 23-AUG-1999; 99US-0149929P.

PR 23-AUG-1999; 99US-0149929P.

PR 23-AUG-1999; 99US-0149930P.

PR 25-AUG-1999; 99US-0150566P.

PR 26-AUG-1999; 99US-0150884P.

PR 27-AUG-1999; 99US-0151065P.

PR 27-AUG-1999; 99US-0151066P.

PR 30-AUG-1999; 99US-0151080P.

PR 31-AUG-1999; 99US-0151303P.

PR 01-SEP-1999; 99US-0151438P.

PR 01-SEP-1999; 99US-0151930P.

PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0156458P.
 PR 28-SEP-1999; 99US-0156596P.
 PR 29-SEP-1999; 99US-0157117P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 08-OCT-1999; 99US-0157865P.
 PR 09-OCT-1999; 99US-0158029P.
 PR 12-OCT-1999; 99US-0158232P.
 PR 13-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 23-OCT-1999; 99US-0161404P.
 PR 23-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161320P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 3.3%; Score 7; DB 3; Length 215;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 VLTAIFQ 191
 Db 119 VLTAIFQ 125

RESULT 44
 ADB64094
 ID ADB64094 standard; protein; 267 AA.

XX ADB64094;

XX 04-DEC-2003 (first entry)

XX Human protein encoded by clone BRWH20183170.

XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
 cell regeneration; membrane protein; signal transduction-related protein;
 transcription-related protein; osteoporosis; neurological disease;
 cancer; tumour.

XX Homo sapiens.

OS Homo sapiens.

XX EPI1308459-A2.

PN EPI1308459-A2.

XX 07-MAY-2003.
 XX 28-MAR-2002; 2002EP-00007401.
 XX 05-NOV-2001; 2001JP-00379298.
 XX 25-JAN-2002; 2002US-00350978.
 XX (HELI-) HELIX RES INST.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 WPI; 2003-450961/43.
 DR N-PSDB; ADB62124.

XX New polynucleotides and polypeptides, useful for developing a diagnostic
 marker or medicines for regulation of their expression and activity, or
 as targets of gene therapy.

Claim 1; Page: 222pp; English.

CC The invention discloses a polynucleotide comprising a sequence selected
 from 1970 fully defined nucleotide sequences which encode novel
 polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 or its partial peptide, an antibody binding to the polypeptide or peptide
 of the polynucleotide, immunologically assaying the polypeptide or
 peptide of the polynucleotide by contacting the polypeptide or peptide
 with the antibody of the encoded protein, and observing the binding
 between the two, a transformant carrying the polynucleotide in an
 expressible manner and an antisense polynucleotide. The oligonucleotide
 is useful as a primer for synthesising the polynucleotide, or as a probe
 for detecting the polynucleotide. The polynucleotides and encoded
 proteins are useful as pharmaceutical agents and many disease-related
 genes may be included in them, for developing a diagnostic marker or
 medicines for regulation of their expression and activity, or as targets
 of gene therapy. The genes are involved in tissue and/or cell
 regeneration. Membrane proteins, signal transduction-related proteins,
 transcription-related proteins, disease-related proteins and genes
 encoding them can be used as indicators for diseases (e.g. osteoporosis,
 neurological diseases, cancer, tumours). The cDNA may be used to regulate
 the activity or expression of the encoded protein to treat diseases. The
 sequence presented is a protein of the invention. Note: Some of the
 sequence data for this patent is not represented in the printed
 specification, but is based on sequence information supplied by the
 European Patent Office.

SQ Sequence 267 AA;

Query Match 3.3%; Score 7; DB 7; Length 267;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATAAI 40
 Db 45 DDATAAI 51

RESULT 45

ADB85214

ID ADB85214 standard; protein; 295 AA.

XX ADB85214;

XX 04-DEC-2003 (first entry)

DT Rat chimaerin SEQ ID NO:95.

XX rat; streptozocin; kinase; phosphatase; ion channel protein; receptor;

KW transporter; G-protein coupled receptor; GPCR; DNA-binding proteins;

KW protease; enzyme; analgesic; gene therapy; pain; diabetes.

XX OS Rattus norvegicus.
 XX PN EP1284297-A2.
 XX PD 19-FEB-2003.
 XX PF 26-JUL-2002; 2002EP-00255228.
 XX PR 27-JUL-2001; 2001GB-00018354.
 XX PR 07-FEB-2002; 2002GB-00002880.
 XX PA (WARN) WARNER LAMBERT CO.
 XX PI Brooksbank RA, Dixon AK, Lee K, Pinnock RD;
 XX DR N-PSDB; ADB85215.
 XX DR WPI; 2003-364994/35.
 XX DR N-PSDB; ADB85215.
 XX PT Use of gene sequence that is down-regulated in response to streptozocin-
 PT induced diabetes, vector, host cell, animal, polypeptide and antibody, in
 PT screening of compounds for treating or diagnosing pain.
 XX PS Disclosure; Page 162-163; 256pp; English.
 XX CC The invention relates to a novel isolated gene sequence that is down-
 CC regulated in the spinal cord in response to streptozocin-induced
 CC diabetes, or comprising, hybridising or having at least 80% sequence
 CC identity to a sequence whose expression products are kinases,
 CC phosphatases, ion channel proteins, receptors, transporters, G-protein
 CC coupled receptor proteins, DNA-binding proteins, proteases or enzymes,
 CC given in the specification. A gene of the invention has analgesic
 CC activity, and may have a use in gene therapy. The gene sequences, vector,
 CC host cell, animal, polypeptide and antibody are useful for screening of
 CC compounds for diagnosing or treating pain. The kits are useful for
 CC simultaneous, separate or sequential detecting and/or quantifying down-
 CC regulation of a gene sequence in the spinal cord of a mammal in response
 CC to streptozocin-induced diabetes. The compound or pharmaceutical
 CC composition is useful as a medicament for treating or diagnosing pain.
 CC The present sequence represents a protein encoded by a gene of the
 CC invention.
 XX SQ Sequence 295 AA;
 Query Match 3.3%; Score 7; DB 7; Length 295;
 Best Local Similarity 100.0%; Pred.No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 IEARGLK 82
 DB 128 IEARGLK 134
 RESULT 46
 AAE18216
 ID AAE18216 standard; protein; 300 AA.
 XX AC AAE18216;
 XX DT 07-NOV-2000 (first entry)
 XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:73.
 XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX OS Plasmodium falciparum.
 XX PN WO200025728-A2.
 XX PD 11-MAY-2000.
 XX PF 05-NOV-1999; 99WO-US026796.

XX OS 05-NOV-1998; 98US-0107131P.
 XX PR (HOFF/) HOFFMAN S.
 XX PA (CARU/) CARUCCI D.
 XX PA (GARD/) GARDNER M.
 XX PA (VENT/) VENTER J C.
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX DR WPI; 2000-365347/31.
 XX DR Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.
 XX PS Disclosure; Page 171-172; 577pp; English.
 XX CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II); (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection. or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX SQ Sequence 300 AA;
 Query Match 3.3%; Score 7; DB 3; Length 300;
 Best Local Similarity 100.0%; Pred.No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 207 KRELEKN 213
 DB 214 KRELEKN 220
 RESULT 47
 ADB09304
 ID ADB09304 standard; protein; 343 AA.
 XX AC ADB09304;
 XX DT 20-NOV-2003 (first entry)
 XX DE Alloiococcus otitis antigenic protein SEQ ID NO:3244.
 XX KW Alloiococcus otitis; antigenic protein; immunogenic; immunisation;
 XX KW gene therapy; Gram-positive bacterium; infection.
 XX OS Alloiococcus otitis.
 XX PN WO2003048304-A2.
 XX PD 12-JUN-2003.
 XX PF 25-NOV-2002; 2002WO-US036123.
 XX PR 29-NOV-2001; 2001US-0333777P.

PR 18-NOV-2002; 2002US-0426742P.
PA (AMHP) WYETH HOLDINGS CORP.
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX WPI; 2003-505284/47.
XX N-PSDB; ADB09303.
XX New Alloicoccus otitidis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX Claim 33; SEQ ID NO 3244; 1019pp; English.
XX The present invention describes an isolated polynucleotide (I) of
CC Alloicoccus otitidis genomic DNA, which encodes an antigenic protein.
CC Alloicoccus otitidis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
CC expression vector comprising the novel isolated polynucleotide (I), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloicoccus otitidis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloicoccus
CC otitidis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (I) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloicoccus
CC otitidis. The present sequence represents an Alloicoccus otitidis
CC antigen protein from the present invention.
XX
SQ Sequence 343 AA;
Query Match 3.3%; Score 7; DB 6; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 FVVALSL 144
Db 209 FVVALSL 215
|||||
RESULT 48
AAR33439
ID AAR33439 standard; protein; 354 AA.
XX
AC AAR33439;
XX
XX 17-DEC-2001 (revised)
DT 06-JUL-1993 (first entry)
XX
XX Ornithine cyclodeaminase C58 from Ti plasmid pTic58.
XX mu-crystallins; drug targeting; nervous acting drugs; CNS; neural;
KW neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
KW neuromuscular junctions; memory agents; Alzheimers disease;
KW CNS depressants; CNS stimulators; tranquilizers; muscle relaxants;
KW antispasmodics; analgesics; anesthetics; anticonvulsants;
KW antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
KW hypnotics.
XX Agrobacterium tumefaciens.
OS

XX USN7844304-N.
PN
XX 01-JAN-1993.
PD
XX 28-FEB-1992; 92US-00844304.
PF
XX 28-FEB-1992; 92US-00844304.
PR
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
PA
XX Wistow G, Kim R;
PI
XX WPI; 1993-093573/11.
DR
XX New mu-crystalline proteins - having ornithine cyclo-deaminase activity,
PT used in diagnosis and treatment of disorders in ornithine metabolism.
PT Disclosure; Page 34; 60pp; English.
XX This sequence represents ornithine cyclodeaminase (OCD) from
CC Agrobacterium Ti plasmid pTic58. It shows approximately 30% homology with
CC the kangaroo eye lens protein mu-crystallin. (Note: Revised entry
CC submitted to correct the patent number format of US Government-owned NTIS
CC applications to prevent clashes with ongoing US granted patent numbers.
CC For further information please visit the Derwent web site at
CC www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ Sequence 354 AA;
Query Match 3.3%; Score 7; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 179 FGVLSDV 185
Db 96 FGVLSDV 102
|||||
RESULT 49
AAR33440
ID AAR33440 standard; protein; 357 AA.
XX
AC AAR33440;
XX
XX 17-DEC-2001 (revised)
DT 06-JUL-1993 (first entry)
XX
XX Ornithine cyclodeaminase Ach5 from Ti plasmid Ach5.
XX mu-crystallins; drug targeting; nervous acting drugs; CNS; neural;
KW neuronal; neurotransmitter agents; neuromuscular agents; NMJ;
KW neuromuscular junctions; memory agents; Alzheimers disease;
KW CNS depressants; CNS stimulators; tranquilizers; muscle relaxants;
KW antispasmodics; analgesics; anesthetics; anticonvulsants;
KW antiepileptic agents; antianxiety agents; hallucinogens; sedatives;
KW hypnotics.
XX Agrobacterium tumefaciens.
OS
XX USN7844304-N.
PN
XX 01-JAN-1993.
PD
XX 28-FEB-1992; 92US-00844304.
PF
XX 28-FEB-1992; 92US-00844304.
PR
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
PA
XX Wistow G, Kim R;
PI
XX WPI; 1993-093573/11.
DR

XX New mu-crystalline proteins - having ornithine cyclo-deaminase activity,
 PT used in diagnosis and treatment of disorders in ornithine metabolism.
 XX
 XX Disclosure; Page 34; 60pp; English.
 XX
 XX This sequence represents ornithine cyclodeaminase (OCD) from
 CC Agrobacterium Ti plasmid Ach5. It shows approximately 30% homology with
 CC the kangaroo eye lens protein mu-crystallin. (Note: Revised entry
 CC submitted to correct the patent number format of US Government-owned NTIS
 CC applications to prevent clashes with ongoing US granted patent numbers.
 CC For further information please visit the Derwent web site at
 CC www.derwent.com/dwpi/updates/ntis_us.html.)
 XX
 XX Sequence 357 AA;

Query Match 3.3%; Score 7; DB 2; Length 357;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 FGVLSDV 185
 |||||
 DB 96 FGVLSDV 102

RESULT 50
 ABU35415
 ID ABU35415 standard; protein; 368 AA.

XX ABU35415;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #20942.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Moraxella catarrhalis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

XX N-PSDB; ACA39285.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 63339; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs; or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 368 AA;

Query Match 3.3%; Score 7; DB 6; Length 368;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 PTHRVIS 133
 |||||
 DB 79 PTHRVIS 85

RESULT 51

ABU40439
 ID ABU40439 standard; protein; 372 AA.

XX ABU40439;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #25966.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX *Pseudomonas putida*.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

XX N-PSDB; ACA44309.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 68363; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 372 AA;

Query Match 3.3%; Score 7; DB 6; Length 372;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 TKVLAPA 205

DB 268 TKVLAPA 274

RESULT 53

ABP69823

ID ABP69823 standard; protein; 383 AA.

XX

AC ABP69823;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human polypeptide SEQ ID NO 1870.

XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;

KW cell-proliferative disorder; neurodegenerative disease; bacterial;

KW Parkinson's disease; Alzheimer's disease; autoimmune disease;

KW multiple sclerosis; diabetes; genetic disorder; wound; infection;

KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;

KW antiParkinsonian; antidiabetic; immunosuppressive; dermatological;

KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

KW antiarthritic.

XX

OS Homo sapiens.

XX

PN WO200270539-A2.

XX

PD 12-SEP-2002.

XX

PF 05-MAR-2002; 2002WO-US005095.

XX

PR 05-MAR-2001; 2001US-00799451.

XX (HYSE-) HYSEQ INC.

XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX

DR WPI; 2002-759812/82.

DR N-PSDB; ABZ12040.

XX

PT New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative, or platelet

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet

PT or coagulation disorders.

XX

PS Claim 9; SEQ ID NO 1870; 1012pp + Sequence Listing; English.

XX

CC The invention relates to an isolated polynucleotide (I) comprising a

CC nucleotide sequence selected from any of 948 sequences (ABZ11119-

CC ABZ12086) or their mature protein coding portion, active domain coding

CC protein or complementary sequences. The polynucleotides are useful for

CC identifying expressed genes or for physical mapping of human genome. The

CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight

CC markers, as a food supplement, for generating antibodies, in medical

CC imaging, screening and diagnostic assays and for treating cell-

CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's

CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,

CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,

CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver

CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),

CC arthritis, etc. Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 383 AA;

Query Match 3.3%; Score 7; DB 5; Length 383;

Best Local Similarity 100.0%; Pred. No. 3.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATAAI 40

DB 289 DDATAAI 295

RESULT 53

ADB09306

ID ADB09306 standard; protein; 425 AA.

XX

AC ADB09306;

XX

DT 20-NOV-2003 (first entry)

XX

DE Alloiooccus otitis antigenic protein SEQ ID NO:3246.

XX

KW Alloiooccus otitis; antigenic protein; immunogenic; immunisation;

KW gene therapy; Gram-positive bacterium; infection.

XX

OS Alloiooccus otitis.

XX

PN WO2003048304-A2.

XX

PD 12-JUN-2003.

XX

PF 25-NOV-2002; 2002WO-US036123.

XX

PR 29-NOV-2001; 2001US-0333777P.

PR 18-NOV-2002; 2002US-0426742P.

XX

PA (AMHP) WIETH HOLDINGS CORP.

XX

PI Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;

XX

DR WPI; 2003-505284/47.
 DR N-PSDB; ADB09305.
 XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.
 XX
 XX Claim 33; SEQ ID NO 3246; 1019pp; English.
 XX
 XX The present invention describes an isolated polynucleotide (I) of
 CC Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.
 CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC expression vector comprising the novel isolated polynucleotide (I), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Alloiooccus otitidis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Alloiooccus
 CC otitidis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting Alloiooccus
 CC otitidis. The present sequence represents an Alloiooccus otitidis
 CC antigen protein from the present invention.
 XX
 SQ Sequence 425 AA;
 Query Match 3.3%; Score 7; DB 6; Length 425;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02; Mismatches 0; Indels 0;
 Matches 7; Conservative 0; Gaps 0;
 QY 138 FVVALSL 144
 Db 291 FVVALSL 297
 |||||
 RESULT 54
 ABU23589
 ID ABU23589 standard; protein; 430 AA.
 XX
 AC ABU23589;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #9116.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Clostridium acetobutylicum.
 XX
 FN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX

PA (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA27459.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 25; SEQ ID NO 51513; 1766pp; English.
 XX
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 430 AA;
 Query Match 3.3%; Score 7; DB 6; Length 430;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02; Mismatches 0; Indels 0;
 Matches 7; Conservative 0; Gaps 0;
 QY 22 YDKITEE 28
 Db 192 YDKITEE 198
 |||||
 RESULT 55
 ABB54464
 ID ABB54464 standard; protein; 434 AA.
 XX
 AC ABB54464;
 XX
 DT 29-AUG-2003 (revised)
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein pbuX.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 OS Lactococcus lactis; IL1403.
 XX
 FN FR2807446-A1.

PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M., Li PWD, Myers EW;
 DR WPI; 2001-658660/75.
 DR N-PSDB; ABL07638.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PS Disclosure; SEQ ID NO 17397; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB161176-AB161177), expressed DNA
 CC sequences (AB161176-AB161177) and the encoded proteins (AB161177-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 460 AA;
 Query Match 3.3%; Score 7; DB 4; Length 460;
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 KFLJIAA 9
 Db 326 KFLJIAA 332
 RESULT 58
 ABR58660
 ID ABR58660 standard; protein; 466 AA.
 AC ABR58660;
 DT 09-JUL-2003 (first entry)
 DE Human cancer related protein SEQ ID NO:317.
 XX Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
 KW heart disease; atherosclerosis; endometriosis.
 XX Homo sapiens.
 OS WO2003025138-A2.
 PN 27-MAR-2003.
 PD 17-SEP-2002; 2002WO-US029560.
 PF 17-SEP-2001; 2001US-0323469P.
 PR 20-SEP-2001; 2001US-0323887P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 08-FEB-2002; 2002US-0358145P.
 PR 08-FEB-2002; 2002US-0355257P.
 PR 12-APR-2002; 2002US-0372246P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;
 XX Zlotnik A;
 PI WPI; 2003-354600/33.
 DR N-PSDB; ACC72811.
 XX
 XX New genes that are up-regulated or down-regulated in cancers, useful as
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
 PT therapeutic targets for screening drugs for treating these diseases.
 XX Claim 12; Page 756; 767pp; English.
 PS The present invention describes an isolated nucleic acid molecule, which
 CC comprises the sequence of any of the genes that are up-regulated or down-
 CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
 CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer
 CC related gene nucleotide sequences which encode the proteins given in
 CC ABR58521 to ABR58709. Also described: (1) determining the presence or
 CC absence of a pathological cell in a patient; (2) an expression vector
 CC comprising a nucleic acid molecule described above; (3) a host cell
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
 CC of (4); (6) specifically targeting a compound to a pathological cell in a
 CC patient by administering to the patient the antibody above; and (7) a
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or
 CC therapeutic targets. In particular, the nucleic acid is useful for
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
 CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies
 XX
 SQ Sequence 466 AA;
 Query Match 3.3%; Score 7; DB 6; Length 466;
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 IEARGLK 82
 Db 299 IEARGLK 305
 RESULT 59
 ABU25000
 ID ABU25000 standard; protein; 468 AA.
 AC ABU25000;
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by prokaryotic essential gene #10527.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Clostridium difficile.
 OS WO200277183-A2.
 PN 03-OCT-2002.
 PD 21-MAR-2002; 2002WO-US009107.
 PF 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 PI WPI; 2003-029926/02.
 DR N-PSDB; ACA28870.
 XX

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 52924; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 468 AA;

Query Match 3.3%; Score 7; DB 6; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 DLAAYKLG 123

Db 41 DLAAYKLG 47

RESULT 60

ADE59897

ID ADE59897 standard; protein; 468 AA.

AC ADE59897;

XX 29-JAN-2004 (first entry)

XX Human Protein P52757, SEQ ID NO 5793.

XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; Chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.

XX Homo sapiens.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; P52757.

PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page: 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 468 AA;

Query Match 3.3%; Score 7; DB 7; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 IEARGLK 82

Db 301 IEARGLK 307

RESULT 61

ABB92783

ID ABB92783 standard; protein; 474 AA.

XX ABB92783;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1994.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP009892.

XX Example 1; Fig 4; 62pp; English.

PS The invention relates to an isolated L-arabinose isomerase active enzyme

CC (1) derived from a *Thermoanaerobacter* sp. or its active fragment. The

CC enzyme is useful for converting an aldose into a ketose which involves

CC contacting the aldose with (1), and keeping the reaction under conditions

CC where at least 1% by weight of the aldose is converted. The aldose is

CC chosen from L-arabinose, D-galactose and D-fucose, and the conversion

CC reaction takes place at 60plusOC. The method further comprises the step

CC of hydrolyzing lactose to obtain D-galactose by use of a lactase -active

CC enzyme, where D-galactose is converted to D-tagatose. The lactase is

CC immobilized and the isomerase enzyme preparation is immobilized, and

CC where the lactose hydrolysis and the aldose isomerization takes place in

CC the same reactor. The lactase-active enzyme is preferably beta-

CC glycosidase. Sequences ABP71431-441 represent L-arabinose isomerase

CC (araA) enzymes from various species used for comparison studies with a T.

CC matrahniil araA enzyme

XX Sequence 498 AA;

SEQ

Query Match 3.3%; Score 7; DB 6; Length 498;

Best Local Similarity 100.0%; Pred. No. 3.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGGLSIL 175

Db 90 IGGLSIL 96

|||||

RESULT 64

ABB62629

ID ABB62629 standard; protein; 511 AA.

XX

AC ABB62629;

XX

XX 26-MAR-2002 (first entry)

DT

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 14679.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

KW

OS Drosophila melanogaster.

OS

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US009231.

PF

XX

PR 23-MAR-2000; 2000US-0191637P.

PR

PR 11-JUL-2000; 2000US-00614150.

XX

XX (PEKE) PE CORP NY.

PA

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

PI

XX

XX WPI; 2001-656860/75.

DR

DR N-PSDB; ABL06732.

XX

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell

PT interactions.

PT

XX

XX Disclosure; SEQ ID NO 14679; 21pp + Sequence Listing; English.

PS

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SEQ Sequence 511 AA;

Query Match 3.3%; Score 7; DB 4; Length 511;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 KVLAPAF 206

Db 129 KVLAPAF 135

|||||

RESULT 65

ADB09310

ID ADB09310 standard; protein; 521 AA.

XX

AC ADB09310;

XX

XX 20-NOV-2003 (first entry)

DT

XX

DE *Alloicoccus otitis* antigenic protein SEQ ID NO:3250.

XX

XX *Alloicoccus otitis*; antigenic protein; immunogenic; immunisation;

KW gene therapy; Gram-positive bacterium; infection.

KW

OS *Alloicoccus otitis*.

OS

XX

PN WO2003048304-A2.

XX

PD 12-JUN-2003.

XX

XX 25-NOV-2002; 2002WO-US036123.

PF

XX

PR 29-NOV-2001; 2001US-0333777P.

PR

PR 18-NOV-2002; 2002US-0426742P.

XX

XX (AMHP) WYETH HOLDINGS CORP.

PA

XX

PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

PI

XX

XX WPI; 2003-505284/47.

DR

DR N-PSDB; ADB09309.

XX

XX

PT New *Alloicoccus otitis* polynucleotides and polypeptides, useful for

PT treating and diagnosing diseases, drug screening assays and monitoring of

PT effects during drug clinical trials.

PT

XX

XX Claim 33; SEQ ID NO 3250; 1019pp; English.

PS

XX

XX The present invention describes an isolated polynucleotide (I) of

CC *Alloicoccus otitis* genomic DNA, which encodes an antigenic protein.

CC *Alloicoccus otitis* is a Gram-positive bacterium. Also described: (1) an

CC isolated polypeptide that is encoded by the polynucleotide (I); (2) an

CC expression vector comprising the novel isolated polynucleotide (I), its

CC complement, degenerate variant or fragment; (3) a genetically engineered

CC host cell, transfected, transformed or infected with the vector of (2);

CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic

CC composition comprising the polypeptide, its complement, biological

CC equivalent or fragment, or the polynucleotide that is comprised in the

CC expression vector; (6) a pharmaceutical composition comprising the

CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array

CC of the polypeptides of (1), their biological equivalent or fragment; (8)

CC immunising against *Alloicoccus otitis* by administering to a host the

CC immunogenic composition; (9) detecting and/or identifying *Alloicoccus*

CC *otitis* in the biological sample; (10) a kit comprising a container

CC containing the novel polynucleotide, its degenerate variant or fragment,

CC or the antibody of (4); and (11) producing a polypeptide by culturing the

CC genetically engineered host cell under conditions suitable to produce the

CC polypeptide from the culture. (I) can be used in gene therapy. The

CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting *Alloccoccus*
 CC otitidis. The present sequence represents an *Alloccoccus* otitidis
 CC antigen protein from the present invention.

XX
 XX Sequence 521 AA;

Query Match 3.3%; Score 7; DB 6; Length 521;

Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 FVVALSL 144

Db 387 FVVALSL 393

RESULT 66

ABJ39024

ID ABJ39024 standard; protein; 535 AA.

XX AC ABJ39024;

DT 17-OCT-2003 (first entry)

DE Mouse class II cytokine receptor SJ2368.

XX Class II cytokine receptor; SJ2368; autoimmune; inflammatory; cytostatic;
 KW allergic disease; septicemia; tumour; immunosuppressive; antiallergic;
 KW antiinflammatory; mouse.

XX Mus sp.

OS WO2003031620-A1.

PN 17-APR-2003.

XX 02-OCT-2002; 2002WO-JP010280.

PF 02-OCT-2001; 2001JP-00306851.

PR 12-JUL-2002; 2002JP-00204395.

XX (MOCH) MOCHIDA PHARM CO LTD.

PA (KAZU) KAZUSA DNA RES INST.

PI Ohara O, Nagase T, Katou Y, Takahashi T, Ohkawa K, Shirakawa K;

DR WPI; 2003-381719/36.

DR N-PSDB; ABT43585.

XX Class II cytokine receptor SJ2368 and regulators of its activity and
 PT expression for treatment and diagnosis of autoimmune, inflammatory and
 PT allergic diseases and tumours.

PS Claim 2; Page 159-162; 188pp; Japanese.

XX This invention relates to the class II cytokine receptor gene SJ2368 and
 CC the encoded protein, derived from either a mouse or human origin.

CC Agonists or antagonists of the cytokine receptor SJ2368 can be used for
 CC the treatment and diagnosis of autoimmune, inflammatory and allergic
 CC diseases, as well as for treating the effects of septicemia and for
 CC tumours. Accordingly, they can be described as having immunosuppressive,
 CC antiinflammatory, antiallergic and/or cytostatic activity. This
 CC polypeptide sequence is the mouse class II cytokine receptor SJ2368 of
 CC the invention

XX Sequence 535 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLIAAVA 11

Db 237 LLIAAVA 243

RESULT 67

ADC17327

ID ADC17327 standard; protein; 535 AA.

XX AC ADC17327;

DT 18-DEC-2003 (first entry)

XX Mouse IFN-lambdaR1 (CRF2-12) protein.

XX mouse; IFN-lambdaR1; CRF2-12; interferon; IFN; alpha/beta independent;
 KW receptor; antiviral; IFN-lambda; IFN-lambda2; IFN-lambda3; IL-10R2;
 KW CRF2-4; cytostatic; virucidal; anti-microbial; hepatotropic;
 KW immunosuppressive; gene therapy; infection; inflammation; anti-cancer;
 KW autoimmune; hyperproliferative; cardiovascular; neurological disorder;
 KW murine.

XX Mus musculus.

PN WO2003066002-A2.

XX 14-AUG-2003.

XX 10-FEB-2003; 2003WO-US003942.

PR 08-FEB-2002; 2002US-0355196P.

PR 15-OCT-2002; 2002US-0418474P.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Kotenko S, Gallagher G;

WPI; 2003-731476/69.

DR N-PSDB; ADC17326.

XX Isolated nucleic acid molecule comprising a sequence that encodes the
 PT CRF2-12 polypeptide, useful for treating leukemias, myeloma, viral
 PT diseases, hepatitis, and autoimmune diseases.

PS Disclosure; Fig 13; 325pp; English.

XX This invention relates to a novel interferon (IFN) alpha/beta independent
 CC ligand receptor system that provides antiviral protection. Specifically,
 CC it refers to three genes located on chromosome 19, designated IFN-
 CC lambda1, IFN-lambda2 and IFN-lambda3 that are expressed upon viral
 CC infection. These IFN-lambda proteins activate a specific receptor complex
 CC for signalling and induction of various biological activities, it
 CC consists of a novel receptor known as IFN-lambdaR1 (or CRF2-12, a class
 CC II cytokine receptor family member) and a second subunit IL-10R2 (or CRF2-
 CC -4). As such, the IFN-lambda compositions of the present invention can be
 CC described as cytostatic, virucidal, anti-microbial, hepatotropic and
 CC immunosuppressive. Accordingly, via gene therapy routes for example, they
 CC can be used to treat various infections and inflammations, as well as
 CC being useful for anti-cancer therapy, autoimmune, hyperproliferative,
 CC cardiovascular and neurological disorders. This polypeptide sequence is
 CC the mouse CRF2-12 protein, homologous to the human CRF2-12 receptor of
 CC the invention.

XX Sequence 535 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLIAAVA 11

Db 237 LLIAAVA 243

```

RESULT 69
ABG25418
ID ABG25418 standard; protein; 581 AA.
XX AC ABG25418;
XX AC ABG25418;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #25409.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN W0200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS89605.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 55777; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activities. The
XX CC polypeptide and polynucleotide sequences have application in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. NOTE: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 581 AA;

Query Match 3.3%; Score 7; DB 4; Length 581;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 DDAAIAI 40
Db 289 DDAAIAI 295
|||||
|||||

RESULT 69
ABG97496
ID ABG97496 standard; protein; 581 AA.
XX AC ABG97496;
XX AC ABG97496;
XX DT 16-DEC-2002 (first entry)
XX DE Human NOVX16 protein.
XX KW Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction;
XX KW Hodgkin disease; Von Hippel-Lindau syndrome; Alzheimer's disease; stroke;
XX KW tuberous sclerosis; hypercalcaemia; Parkinson's disease; depression;
XX KW Huntington's disease; cerebral palsy; epilepsy; Lesch-Nyhan syndrome;
XX KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain;
XX KW obesity; Crohn's disease; osteoporosis; inflammatory bowel disease;
XX KW infertility; inflammatory bowel disease; atherosclerosis; hypertension;
XX KW scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease;
XX KW asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis;
XX KW bacterial infection; parasitic infection; graft-versus-host disease;
XX KW cell differentiation; cell proliferation; haematopoiesis; wound healing;
XX KW angiogenesis.
XX OS Homo sapiens.
XX PN W0200272770-A2.
XX PD 19-SEP-2002.
XX PF 08-MAR-2002; 2002WO-US007283.
XX PR 08-MAR-2001; 2001US-0274281P.
XX PR 09-MAR-2001; 2001US-0274849P.
XX PR 12-MAR-2001; 2001US-0275235P.
XX PR 13-MAR-2001; 2001US-0275579P.
XX PR 13-MAR-2001; 2001US-0275601P.
XX PR 14-MAR-2001; 2001US-0276000P.
XX PR 20-MAR-2001; 2001US-0277239P.
XX PR 20-MAR-2001; 2001US-0277327P.
XX PR 20-MAR-2001; 2001US-0277338P.
XX PR 21-MAR-2001; 2001US-0277791P.
XX PR 22-MAR-2001; 2001US-0277833P.
XX PR 23-MAR-2001; 2001US-0278152P.
XX PR 26-MAR-2001; 2001US-0278894P.
XX PR 27-MAR-2001; 2001US-0279036P.
XX PR 28-MAR-2001; 2001US-0279344P.
XX PR 30-MAR-2001; 2001US-0280233P.
XX PR 02-APR-2001; 2001US-0280802P.
XX PR 02-MAY-2001; 2001US-0288148P.
XX PR 31-MAY-2001; 2001US-0294821P.
XX PR 31-OCT-2001; 2001US-0335302P.
XX PR 04-DEC-2001; 2001US-0338375P.
XX PR 07-MAR-2002; 2002US-00094466.
XX PA (CURA-) CURAGEN CORP.
XX PI Spytek KA, Vernet CA, Tchernev VT, Malvankar UM, Gerlach VL;
XX PI Li L, Zehrsen BD, Patturajan M, Gusev VI, Kekuda R, Pena CEA;
XX PI Zhong M, Gangolli EA, Taupier RJ;
XX WPI; 2002-713508/77.
XX DR N-PSDB; ABS78740.
XX PT New NOVX polypeptides and polynucleotides, useful for preventing,
XX PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple
XX PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or
XX PT Parkinson's disease.
XX PS Claim 1; Page 137; 266pp; English.
XX CC The present invention relates to a new polypeptide (NOVX). The NOVX
XX CC polypeptide, nucleic acid and antibody are useful in the manufacture of a
XX CC medicament for treating a syndrome associated with a human disease,
XX CC preferably a NOVX-associated disorder. The NOVX nucleic acids,

```

CC polypeptides and antibodies are useful for treating, preventing or
 CC diagnosing diseases such as cancers, Hodgkin disease, Von Hippel-Lindau
 CC syndrome, Alzheimer's disease, stroke, tuberculous sclerosis, cerebral
 CC hypercalcaemia, Parkinson's disease, Huntington's disease, ataxia-
 CC palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-
 CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,
 CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,
 CC infertility, inflammatory bowel disease, atherosclerosis, hypertension,
 CC scleroderma, haemophilia, diabetes, pancreatitis, autoimmune disease,
 CC asthma, arthritis, immunodeficiencies, HIV, viral, bacterial or parasitic
 CC infections, or graft-versus-host disease. The nucleic acids and
 CC polypeptides may also be used as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOVX substances for use in therapeutic or
 CC diagnostic methods. The nucleic acids are further used as hybridisation
 CC probes, in chromosome mapping, tissue typing, preventive medicine, and
 CC pharmacogenomics. The present amino acid sequence represents a human NOVX
 CC protein of the invention
 XX
 CC Sequence 581 AA;

Query Match 3.3%; Score 7; DB 5; Length 581;

Best Local Similarity 100.0%; Pred. No. 4.4e+02; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 DDATAAI 40

DB 289 DDATAAI 295

|||||||

RESULT 70

ABU20412

ID ABU20412 standard; protein; 625 AA.

AC ABU20412;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #5939.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Bacteroides fragilis.

XX WO20027183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI WPI; 2003-029926/02.

XX N-PSDB; ACA24282.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 48336; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX

CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 CC Sequence 625 AA;

Query Match 3.3%; Score 7; DB 6; Length 625;

Best Local Similarity 100.0%; Pred. No. 4.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 PAFKREL 210

DB 525 PAFKREL 531

|||||||

RESULT 71

ADB80157

ID ADB80157 standard; protein; 657 AA.

XX ADB80157;

AC ADB80157;

XX 04-DEC-2003 (first entry)

DT Mycobacterium tuberculosis nutrient starvation-inducible protein #98.

DE mycobacteria; nutrient starving condition; mycobacterial latency;

XX mycobacterial infection; vaccine.

XX Mycobacterium tuberculosis.

OS WO2003004520-A2.

XX 16-JAN-2003.

XX 04-JUL-2002; 2002WO-CB003052.

XX 04-JUL-2001; 2001GB-00016385.

PR 05-OCT-2001; 2001GB-00023993.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PA James BW, Marsh P, Hampshire T;

PI WPI; 2003-210338/20.

XX N-PSDB; ADB80158.

XX

PT New mycobacterial peptide, useful for the manufacture of a medicament for
 PT treating or preventing, or a diagnostic reagent for identifying,
 PT mycobacterial infection.

XX Claim 2; Page 329-332; 442pp; English.

XX The invention comprises the amino acid and coding sequences of
 CC mycobacterial (Mycobacterium tuberculosis) proteins which are up-
 CC regulated under nutrient starving conditions and maintain mycobacterial
 CC latency. The DNA and protein sequences of the invention are useful for
 CC the treatment, prevention and diagnosis of a mycobacterial infection. The
 CC present amino acid sequence represents a Mycobacterium tuberculosis
 CC protein of the invention.

XX Sequence 657 AA;

Query Match 3.3%; Score 7; DB 7; Length 657;
 Best Local Similarity 100.0%; Pred. No. 5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLIIAAV 10

Db 324 FLIIAAV 330

RESULT 72

AAR38153
 ID AAR38153 standard; protein; 752 AA.

AC AAR38153;

DT 24-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 13-OCT-1993 (first entry)

DE Acetobacter diguanylate phosphodiesterase PDEA2.

XW Cyclic diguanylate; diguanylate phosphodiesterase; diguanylate cyclase;
 XW cellulose production; cdg2 operon.

OS Gluconacetobacter xylinus.

Key Location/Qualifiers

FT Misc-difference 77 /note= "Val deduced from GG"

FT Misc-difference 212 /note= "Met deduced from TAG"

FT Misc-difference 215 /note= "Arg deduced from GC"

FT Misc-difference 232 /note= "Met deduced from ATC"

FT Misc-difference 233 /note= "Leu deduced from GTG"

FT Misc-difference 247 /note= "Gly deduced from GC"

FT Misc-difference 271 /note= "Gln deduced from CAGG"

FT Misc-difference 275 /note= "Val deduced from GCTG"

FT Misc-difference 388 /note= "Asp deduced from GAA"

FT Misc-difference 538 /note= "Ala deduced from CC"

XX WO9311244-A1.

EN 10-JUN-1993.

XX 14-OCT-1992; 92WO-US008756.

XX 29-NOV-1991; 91US-00800218.

XX (WEYE) WEYERHAEUSER CO.

XX Tal R, Gelfand DH, Calhoon RD, Ben-Bassat A, Benziman M, Wong HC;
 PI WPI; 1993-197062/24.
 DR N-PSDB; AAQ43661.

XX Polynucleotide sequence from acetobacter CDG operon - encodes cyclic
 PT di:guanosine mono:phosphate degradation enzymes e.g. 3-phosphodiesterase
 PT isozyme.

XX Claim 5; Page 80-83; 98pp; English.

CC The amino acid sequence of protein PDEA2 was deduced from the 1st. open
 CC reading frame of the cdg2 operon. The protein is a diguanylate
 CC phosphodiesterase A, i.e. it enzymatically cleaves a single
 CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG. See also
 CC AAR38154. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24-
 CC OCT-2003 to Standardise OS field)

XX Sequence 752 AA;

Query Match 3.3%; Score 7; DB 2; Length 752;
 Best Local Similarity 100.0%; Pred. No. 5.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEQSET 45

Db 304 AIEQSET 310

RESULT 73

AB93713

ID ABB93713 standard; protein; 791 AA.

AC ABB93713;

XX 31-MAY-2002 (first entry)

DT Herbicidally active polypeptide SEQ ID NO 2924.

DE Herbicidally active polypeptide SEQ ID NO 2924.

XX Herbicidally active polypeptide; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP009892.

XX 28-AUG-2001; 2001WO-EP009892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidner M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.

XX Claim 5; SEQ ID NO 2924; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as

CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation or screening for homologous nucleic acids
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1001 AA;

Query Match 3.3%; Score 7; DB 6; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 7.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AVSADPI 20
 DB 181 AVSADPI 187

RESULT 76

ID ADC01332 standard; protein; 1027 AA.

XX ADC01332;

DT 04-DEC-2003 (first entry)

DE Enterohaemorrhagic *E. coli* O157:H7-specific protein SEQ ID NO: 1377.

XX enterohaemorrhagic; anti-bacterial.

XX *Escherichia coli*; O157:H7.

XX JP2002355074-A.

XX 10-DEC-2002.

XX 24-JAN-2002; 2002JP-00015959.

XX 24-JAN-2001; 2001JP-00112010.

XX (UUTS-) UNIV TSUKUBA.

XX WPI; 2003-451640/43.

XX Enterohaemorrhagic *Escherichia coli* O157:H7-specific nucleic acid molecule
 PT and a polypeptide and its use, a polypeptide, a vector and a host cell.

XX Claim 3; SEQ ID NO 1377; 2067pp; Japanese.

XX The invention relates to a novel enterohaemorrhagic *Escherichia coli*
 CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
 CC has anti-bacterial activity. The polypeptide can be used in detection
 CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
 CC genome of Enterohaemorrhagic *E. coli* O157:H7 was determined. The present
 CC sequence represents an *E. coli* O157:H7-specific polypeptide of the
 CC invention.

XX Sequence 1027 AA;

Query Match 3.3%; Score 7; DB 7; Length 1027;
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLIIAAV 10
 DB 135 FLIIAAV 141

RESULT 77

ID ABM15822

XX ABM15822 standard; protein; 1194 AA.

XX AC ABM15822;

DT 26-SEP-2003 (first entry)

XX Mycobacterium tuberculosis mycobacterial antigen protein SEQ ID NO:11.

XX Mycobacterium tuberculosis; mycobacterial; antigen; infection; vaccine;

XX tuberculostatic; mycobacterial peptide; mycobacterial infection.

XX Mycobacterium tuberculosis.

XX WO2003033530-A2.

XX 24-APR-2003.

XX 14-OCT-2002; 2002WO-GB004647.

XX 12-OCT-2001; 2001GB-00024593.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX James B, Bacon J, March P;

XX WPI; 2003-393501/37.

XX N-PSDB; ACF39318.

XX New isolated mycobacterial peptide encoded by a gene that is induced or
 PT up-regulated under high oxygen tension, useful for diagnosing, treating
 PT or preventing a mycobacterial infection.

XX Claim 1; Page 81-86; 392pp; English.

XX The present invention describes an isolated mycobacterial peptide (I), or
 CC its fragment, variant or derivative encoded by a gene whose expression is
 CC induced or up-regulated during culture of a mycobacterium under
 CC continuous culture conditions of a dissolved oxygen tension of at least
 CC 30% air saturation measured at 37 plus degrees Celsius when compared with
 CC a dissolved oxygen tension of up to 10% air saturation measured at 37
 CC plus degrees Celsius. (I) has tuberculostatic activity and can be used in
 CC vaccines. The mycobacterial peptide (I) or its fragment, variant or
 CC derivative, inhibitor, antibody, attenuated mycobacterium, attenuated
 CC microbial carrier, DNA sequence, DNA plasmid, RNA sequence, or RNA vector
 CC from the present invention can be used for manufacturing a medicament for
 CC treating or preventing a mycobacterial infection. The peptide or its
 CC fragment, variant or derivative, the antibody, or a polynucleotide probe
 CC comprising at least 8 nucleotides, where the probe binds to at least a
 CC part of the gene, is useful for manufacturing a diagnostic reagent for
 CC identifying a mycobacterial infection. The present sequence represents a
 CC Mycobacterium tuberculosis mycobacterial antigen, which is used in the
 CC exemplification of the present invention

XX Sequence 1194 AA;

Query Match 3.3%; Score 7; DB 6; Length 1194;
 Best Local Similarity 100.0%; Pred. No. 8.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 IDDAIAA 39
 DB 45 IDDAIAA 51

RESULT 78

ID ADC01365

XX ADC01365 standard; protein; 1400 AA.

XX ADC01365;

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2836 AA;

Query Match 3.3%; Score 7; DB 4; Length 2836;
Best Local Similarity 100.0%; Pred. No. 1.8e-03; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 72 AMENIERA 78
DB 1376 AMENIERA 1382
|||||

RESULT 81
AAB23790
ID AAB23790 standard; peptide; 8 AA.
XX
AC AAB23790;
XX
DT 12-JAN-2001 (first entry)
XX
DE Voltage sensitive cell adhesion molecule synthetic peptide PA22-2.
XX
KW Voltage sensitive cell; VSC; adhesion molecule; binding molecule; neuron;
KW neurite promoting domain; hybrid electrical device; electrical junction;
KW transistor; artificial chemical synapse.
XX
OS Synthetic.
XX
PN WO200051191-A1.
XX
PD 31-AUG-2000.
XX
PF 22-FEB-2000; 2000WO-IL000112.
XX
PR 22-FEB-1999; 99US-0121237P.
XX
PS (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI Yitzchaik S, Shappir J, Spira ME;
XX
DR WPI; 2001-015536/02.
XX
PT A hybrid electrical device with biological components that provide an
PT artificial chemical synapse.
XX
PS Example 6; Page 13; 48pp; English.
XX
CC The present invention describes a hybrid electrical device with
CC biological components that provide an artificial chemical synapse. An
CC electrical junction (I) between one transistor and at least 1 voltage-
CC sensitive cell (VSC) characterised by at least 1 of the features selected
CC from: (i) voltage transfer between the transistor and the VSC is
CC bidirectional; (ii) there is no DC bias between the transistor and the
CC solution containing the VSC; (iii) the VSC is anchored to the external
CC surface of the transistor by binding moieties, optionally through spacers
CC ; and/or (iv) the voltage transfer between the membrane of the VSC and
CC the external surface of the transistor, and between the external surface
CC of the transistor and the membrane of the VSC is mediated by hyper-
CC polarisable chromophores. As an artificial chemical synapse i.e. a
CC junction between a cell (which secretes an agent) and a transistor
CC bearing receptors for the agent. Binding of the agent to the receptor
CC changes an electrical property of the transistor. The present sequence
CC represents a synthetic peptide, designated PA 22-2, which can be used as
CC a VSC adhesion molecule, where the VSC is a neuron, which is used in an
CC example from the present invention
XX

SQ Sequence 8 AA;

Query Match 2.8%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 2 VAVSAD 7
|||||

RESULT 82
ABG70880
ID ABG70880 standard; peptide; 8 AA.
XX
AC ABG70880;
XX
DT 29-AUG-2003 (revised)
DT 29-NOV-2002 (first entry)
XX
DE R. rhodochrous nitrile hydratase beta subunit peptide #2.
XX
KW Nitrile hydratase; beta subunit; enzyme; amide production;
KW cyanide resistance; 2-hydroxy-4-methylthiobutyronitrile;
KW 2-hydroxy-4-methylthiobutyroamide; alpha-hydroxy amide.
XX
OS Rhodococcus rhodochrous; strain J1.
XX
PN WO200270717-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-JP001912.
XX
PR 02-MAR-2001; 2001JP-00059023.
PR 24-JAN-2002; 2002JP-00016222.
XX
PA (DAIL) DAICEL CHEM IND LTD.
XX
PI Nagasawa T, Matsuyama A;
XX
DR WPI; 2002-707100/76.
XX
PT Novel nitrile hydratase which acts on nitrile group of nitrile compound,
PT hydrates nitrile group and converts it to amide group, useful for
PT producing amides using alpha-hydroxy nitrile as starting material.
XX
PS Example 22; Page 55; 96pp; English.
XX
CC The invention relates to a nitrile hydratase which acts on the nitrile
CC group of a nitrile compound, hydrates the nitrile group and converts it
CC to an amide group, and which is resistant to cyanide, or a protein
CC complex consisting of the alpha and beta subunits which act on a nitrile
CC group of a nitrile compound, hydrate the nitrile group and convert it to
CC an amide group, and use 2-hydroxy-4-methylthiobutyronitrile as a
CC substrate to produce 2-hydroxy-4-methylthiobutyroamide. Also included are
CC the polynucleotides encoding the alpha and beta subunits, recombinant
CC vector containing either or both polynucleotides, a transformed host
CC containing the vector, producing an amide (comprising contacting a
CC nitrile compound with the hydratase or a microorganism producing it, or
CC their processed products and recovering the amide produced) and
CC stabilising the activity of nitrile hydratase in the presence of nitrile
CC compounds, comprising contacting the nitrile hydratase with divalent
CC metal cations. The nitrile hydratase, complex or transformed host is
CC useful for producing an amide which involves contacting a nitrile
CC compound such as alpha-hydroxy- nitrile with the enzymatically active
CC material and recovering the amide (preferably alpha-hydroxy amide)
CC produced. The alpha-hydroxynitrile is represented by formula F1, and the
CC product alpha-hydroxy amide is represented by formula F3 (both detailed
CC in the specification). The present sequence represents a peptide from
CC Rhodococcus rhodochrous strain J1 nitrile hydratase beta subunit used to
CC design a PCR primer for isolation of the DNA encoding Rhodococcus sp. Cr4
CC nitrile hydratase. (Updated on 29-AUG-2003 to standardise OS field)
CC

```

XX SQ Sequence 8 AA;
Query Match 2.8%; Score 6; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.1e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NVVNHI 169
DB 2 NVVNHI 7

RESULT 83
AAR70305
ID AAR70305 standard; peptide; 10 AA.
XX AC AAR70305;
XX DT 14-NOV-1995 (first entry)
XX XX Laminin (LM) binding peptide, LM-2.
XX KW Laminin; LM; soluble; receptor; cell-adhesion.
XX OS Synthetic.
XX PN US5399667-A.
XX PD 21-MAR-1995.
XX PF 05-MAR-1993; 93US-00029333.
XX PR 05-MAR-1993; 93US-00029333.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Kosfeld MD, Frazier WA;
XX XX WPI; 1995-130736/17.
XX DT Thrombospondin receptor binding peptide(s) - comprise tri:peptide
PT sequence Val-Val-Met (VVM).
XX PS Disclosure; Fig 4; 25pp; English.
XX CC AAR70302, AAR70303, and AAR70305 are synthetic peptides which are capable
CC of binding the laminin binding domain. These peptides are useful in cell
CC binding assays
XX SQ Sequence 10 AA;
Query Match 2.8%; Score 6; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
DB 4 VAVSAD 9

RESULT 84
ABG67586
ID ABG67586 standard; peptide; 10 AA.
XX AC ABG67586;
XX DT 07-OCT-2002 (first entry)
XX DE Human ADPI tryptic digest peptide #295.
XX KW Human; Alzheimer's disease; AD; brain tissue; ADF; ADPI;
KW Alzheimer's disease-associated feature; neuroprotective;
KW Alzheimer's disease-associated protein isoform; nootropic;

KW ADPI tryptic digest peptide.
XX OS Homo sapiens.
XX PN WO200246767-A2.
XX XX 13-JUN-2002.
XX PF 29-NOV-2001; 2001WO-GB005289.
XX PR 08-DEC-2000; 2000US-0254431P.
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX PI Herath HMC, Parekh RB, Rohlf C;
XX DR WPI; 2002-508575/54.
XX PT Screening, diagnosis or prognosis of Alzheimer's disease in subject,
XX PT comprises detecting Alzheimer disease-associated features or Alzheimer
XX PT disease-associated protein isoforms in brain tissue from the subject.
XX PS Claim 7; Page 66; 427pp; English.
XX CC The present invention relates to methods and compositions for the
XX CC screening, diagnosis or prognosis of Alzheimer's disease (AD) in a
XX CC subject. The method comprises analysing a sample of brain tissue from a
XX CC subject by 2D electrophoresis to generate a 2D array of Alzheimer's
XX CC disease-associated features (ADFs), whose relative abundance correlates
XX CC with the presence, absence, stage or severity of AD and comparing the
XX CC abundance of each feature with the abundance of that chosen feature in
XX CC brain tissue from persons free from AD. The invention also describes in
XX CC Alzheimer's disease-associated protein isoforms (ADPis) detectable in
XX CC brain tissue. The methods and compositions of the invention are useful
XX CC for the screening, diagnosis or prognosis of AD in a subject, for
XX CC determining the stage or severity of AD in a subject, for identifying a
XX CC subject at risk of developing AD, or for monitoring the effect of therapy
XX CC administered to a subject having AD. Antibodies capable of binding to
XX CC ADPis are useful for treating or preventing AD, and for determining the
XX CC efficacy of a given treatment regime. An agent that modulates the
XX CC activity of ADPI is useful in the manufacture of a medicament for the
XX CC treatment or prevention of AD in a subject. ABG67292-ABG68038 represent
XX CC human ADPI tryptic digest peptides
XX SQ Sequence 10 AA;
Query Match 2.8%; Score 6; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 HLLIGV 107
DB 1 HLLIGV 6

RESULT 85
ABU58121
ID ABU58121 standard; peptide; 10 AA.
XX AC ABU58121;
XX DT 11-APR-2003 (first entry)
XX DE Laminin cell binding domain peptide LM22-2.
XX KW Thrombospondin-1; mouse; rat; human; chicken; tumour cell migration;
XX KW cell adhesion; platelet aggregation; wound healing; laminin; TS-1.
XX OS Synthetic.
XX PN US6469138-B1.
XX PD 22-OCT-2002.

```

XX 21-FEB-1995; 95US-00391820.
 XX 05-MAR-1993; 93US-00029333.
 XX (UNIW) UNIV WASHINGTON.
 XX Frazier WA, Kosfeld MD;
 XX WPI; 2003-196751/19.
 XX Novel peptide that binds to thrombospondin 1 receptor, useful as probe
 XX sequence in computer search of all available databases.
 XX Disclosure; Fig 4; 25pp; English.
 XX This invention relates to a novel peptide that binds to the
 XX thrombospondin 1 (TS-1) receptor; Thrombospondin 1 is a protein which
 XX acts to promote cell adhesion and is involved in platelet aggregation,
 XX wound healing and tumour cell migration. Also disclosed in the invention
 XX is Laminin peptides which have been shown to be able to substitute for TS
 XX -1 peptides. Preferably the peptides of the invention share a tripeptide
 XX motif VVM and are useful as probe sequences in a computer search of all
 XX available databases for similar receptor binding motif peptides. The
 XX present sequence represents a thrombospondin-1 (TS-1) or Laminin cell
 XX binding domain peptide of the invention
 XX Sequence 10 AA;
 SQ
 Query Match 2.8%; Score 6; DB 6; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 VAVSAD 18
 DB 4 VAVSAD 9
 RESULT 86
 ADA23705
 ID ADA23705 standard; peptide; 10 AA.
 AC ADA23705;
 XX 20-NOV-2003 (first entry)
 DT Alzheimer's disease-associated protein isoform tryptic peptide #314.
 DE human; Alzheimer's disease; vascular dementia; Lewy body dementia;
 XX schizophrenia; Parkinson's disease; multiple sclerosis; depression;
 KW Alzheimer's disease-associated protein isoform; ADPI.
 XX Homo sapiens.
 OS US2003064411-A1.
 PN 03-APR-2003.
 XX 10-DEC-2001; 2001US-00014340.
 XX 08-DEC-2000; 2000US-0254431P.
 XX (HERA/) HERATH H M A C.
 XX (PARE/) PAREKH R B.
 XX (ROHL/) ROHLFF C.
 XX Herath HMAc, Parekh RB, Rohlf C;
 XX WPI; 2003-540784/51.
 XX Screening, diagnosis or prognosis of Alzheimer's disease in subject,
 PT involves analyzing test sample of brain tissue from subject, and
 PT comparing feature in test sample with that of person(s) free from

PT Alzheimer's disease.
 XX Disclosure; SEQ ID NO 314; 115pp; English.
 XX The invention relates to a method of screening or diagnosing Alzheimer's
 XX disease in a subject. The method is useful for screening, diagnosis or
 XX prognosis of Alzheimer's disease in a subject for determining the stage
 XX of severity of Alzheimer's disease in a subject, for identifying a
 XX subject at risk of developing Alzheimer's disease, or for monitoring the
 XX effect of therapy administered to a subject having Alzheimer's disease.
 XX The method is also useful in treating vascular dementia, Lewy body
 XX dementia, schizophrenia, Parkinson's disease, multiple sclerosis or
 XX depression. The inventive method identifies sensitive and specific
 XX biomarkers for the diagnosis of Alzheimer's disease in living subjects.
 XX It provides therapeutic agents for Alzheimer's disease that works
 XX quickly, potentially, specifically with fewer side effects. The present
 XX sequence represents the amino acid sequence of a Alzheimer's disease-
 XX associated protein isoform tryptic peptide.
 XX Sequence 10 AA;
 SQ
 Query Match 2.8%; Score 6; DB 6; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 102 HLLIGV 107
 DB 1 HLLIGV 6
 RESULT 87
 AAW69442
 ID AAW69442 standard; protein; 11 AA.
 AC AAW69442;
 XX 10-DEC-1998 (first entry)
 DT Rat phospholipase A1 fragment.
 XX Phospholipase A1; phosphatidyl-serine hydrolysis; serine phospholipid;
 KW PLA1; enzyme; rat.
 XX Rattus sp.
 OS Key Location/Qualifiers
 FH Misc-difference 10 /note= "unspecified amino acid"
 FT JP10201479-A.
 XX 04-AUG-1998.
 XX 23-JAN-1997; 97JP-00024269.
 XX 23-JAN-1997; 97JP-00024269.
 XX (TORA) TORAY IND INC.
 XX WPI; 1998-474493/41.
 XX Phospholipase A1 and nucleic acid coding for it - having substrate
 XX specificity against serine phospholipid.
 XX Example 1; Page 5; 13pp; Japanese.
 XX This sequence is a fragment of the rat phospholipase A1 (PLA1) of the
 CC invention. PLA1 is capable of hydrolysing phosphatidyl-serine, and has
 CC substrate specificity for serine phospholipids
 XX Sequence 11 AA;
 SQ
 Query Match 2.8%; Score 6; DB 2; Length 11;

Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 142 LSLEIS 147
| | | | |
Db 1 LSLEIS 6

RESULT 88
AAW73718
ID AAW73718 standard; peptide; 15 AA.
XX
AC AAW73718;
XX
DT 24-MAR-1999 (first entry)
XX
DE M. tuberculosis antigen clone ORF-2 peptide 2-25.
XX
KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX infection.
XX Mycobacterium tuberculosis.
XX WO9853076-A2.
XX
PN 26-NOV-1998.
XX
PD 20-MAY-1998; 98WO-US010514.
XX
PF 20-MAY-1997; 97US-00858998.
XX
PR 05-MAY-1998; 98US-00073009.
XX
XX (CORI-) CORIXA CORP.
XX
PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX WPI; 1999-045315/04.
XX
XX New isolated Mycobacterium tuberculosis antigens - used to develop
PT products for the prevention, treatment and diagnosis of tuberculosis
PT infection.
XX
XX Example 1; Page 64; 104pp; English.
XX
XX This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans);
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
XX Sequence 15 AA;
SQ

Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 SDVLTA 188
| | | | |
Db 8 SDVLTA 13

RESULT 90
AAW73826

Db 3 SDVLTA 8

RESULT 89
AAW73716
ID AAW73716 standard; peptide; 15 AA.
XX
AC AAW73716;
XX
DT 24-MAR-1999 (first entry)
XX
DE M. tuberculosis antigen clone ORF-2 peptide 2-23.
XX
KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
XX infection.
XX Mycobacterium tuberculosis.
XX WO9853076-A2.
XX
PN 26-NOV-1998.
XX
PD 20-MAY-1998; 98WO-US010514.
XX
PF 20-MAY-1997; 97US-00858998.
XX
PR 05-MAY-1998; 98US-00073009.
XX
XX (CORI-) CORIXA CORP.
XX
PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX WPI; 1999-045315/04.
XX
XX New isolated Mycobacterium tuberculosis antigens - used to develop
PT products for the prevention, treatment and diagnosis of tuberculosis
PT infection.
XX
XX Example 1; Page 63; 104pp; English.
XX
XX This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans);
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
XX Sequence 15 AA;
SQ

Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 SDVLTA 188
| | | | |
Db 8 SDVLTA 13

RESULT 90
AAW73826

AAW73826 standard; peptide, 15 AA.
AAW73826;
24-MAR-1999 (first entry)
M. tuberculosis antigen clone ORF-2 peptide 2-23.
Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
infection.
Mycobacterium tuberculosis.
WO9853075-A2.
26-NOV-1998.
20-MAY-1998; 98WO-US010407.
20-MAY-1997; 97US-00859381.
05-MAY-1998; 98US-00073010.
(CORI-) CORIXA CORP.
Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
WPI; 1999-045314/04.
Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
useful for immunisation against M. tuberculosis infection to treat or
prevent tuberculosis, and in diagnosis of tuberculosis.
Example 1; Page 62; 100pp; English.
This sequence represents an immunogenic portion of a Mycobacterium
tuberculosis antigen of the invention. The polypeptides are useful for
immunotherapy to treat or prevent tuberculosis (especially in humans);
e.g. they can be included with an acceptable carrier in pharmaceutical
compositions or included in vaccines, and administered to induce
protective immunity in a patient against M. tuberculosis. Tuberculosis is
a chronic, infectious disease generally caused by M. tuberculosis
infection, and if left untreated typically results in serious
complications and death. Fusion proteins containing the antigen, or DNA
molecules can similarly be included with an acceptable carrier in
pharmaceutical compositions or in vaccines and administered as above. The
polypeptides are also useful for diagnosis of tuberculosis, by contacting
dermal cells with at least one polypeptide and detecting an immune
response (especially induration) on the patient's skin. Inhibiting the
spread of tuberculosis requires vaccination and accurate diagnosis, since
antibiotic therapy may not be effective due to the existence of an
asymptomatic but contagious stage and to patient non-compliance. The
polypeptides overcome concerns of safety and efficacy of current
vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
lack of sensitivity and specificity of existing diagnostic techniques
Sequence 15 AA;
Query Match 2.8%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 SDVLT A 188
| | | | |
Db 8 SDVLT A 13
RESULT 91
AAW73828
ID AAW73828 standard; peptide, 15 AA.
XX
AC AAW73828;
XX
XX 24-MAR-1999 (first entry)

XX DE M. tuberculosis antigen clone ORF-2 peptide 2-25.
XX KW Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW KW infection.
XX OS Mycobacterium tuberculosis.
XX PN WO9853075-A2.
XX PD 26-NOV-1998.
XX PF 20-MAY-1998; 98WO-US010407.
XX PR 20-MAY-1997; 97US-00859381.
XX PR 05-MAY-1998; 98US-00073010.
XX PA (CORI-) CORIXA CORP.
XX PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX DR WPI; 1999-045314/04.
XX PS Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
XX PT useful for immunisation against M. tuberculosis infection to treat or
XX PT prevent tuberculosis, and in diagnosis of tuberculosis.
XX CC Example 1; Page 63; 100pp; English.
XX CC This sequence represents an immunogenic portion of a Mycobacterium
XX CC tuberculosis antigen of the invention. The polypeptides are useful for
XX CC immunotherapy to treat or prevent tuberculosis (especially in humans);
XX CC e.g. they can be included with an acceptable carrier in pharmaceutical
XX CC compositions or included in vaccines, and administered to induce
XX CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
XX CC a chronic, infectious disease generally caused by M. tuberculosis
XX CC infection, and if left untreated typically results in serious
XX CC complications and death. Fusion proteins containing the antigen, or DNA
XX CC molecules can similarly be included with an acceptable carrier in
XX CC pharmaceutical compositions or in vaccines and administered as above. The
XX CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
XX CC dermal cells with at least one polypeptide and detecting an immune
XX CC response (especially induration) on the patient's skin. Inhibiting the
XX CC spread of tuberculosis requires vaccination and accurate diagnosis, since
XX CC antibiotic therapy may not be effective due to the existence of an
XX CC asymptomatic but contagious stage and to patient non-compliance. The
XX CC polypeptides overcome concerns of safety and efficacy of current
XX CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
XX CC lack of sensitivity and specificity of existing diagnostic techniques
XX CC Sequence 15 AA;
XX SQ Query Match 2.8%; Score 6; DB 2; Length 15;
XX SQ Best Local Similarity 100.0%; Pred. No. 1.7e+02;
XX SQ Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 SDVLT A 188
| | | | |
Db 3 SDVLT A 8
RESULT 92
AAU08183
ID AAU08183 standard; peptide, 15 AA.
XX
AC AAU08183;
XX
XX 17-DEC-2001 (first entry)
XX DE Mycobacterium tuberculosis overlapping peptide ORF-2-25.
XX DE Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
KW immunostimulant; mutant; mutein; ORF-2-25.
KW

```

XX OS Mycobacterium tuberculosis.
XX OS Synthetic.
XX FN WO200162893-A2.
XX PD 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US005992.
XX XX 25-FEB-2000; 2000US-0185037P.
XX PR 08-AUG-2000; 2000US-0223828P.
XX XX (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Skeiky Y, Overdale P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX PT response to and inhibiting development of a Mycobacterium infection.
XX PS Example 1; Page 134; 161pp; English.
XX CC The present invention relates to the isolation of Mycobacterium
XX CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
XX CC encoding them. The invention describes compounds and methods for the
XX CC diagnosis of tuberculosis or for inducing protective immunity against
XX CC tuberculosis. The compounds comprise at least one immunogenic portion of
XX CC one or more Mycobacterium proteins and nucleic acid molecules encoding
XX CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
XX CC encoding them can be used in diagnostic kits for the detection of
XX CC Mycobacterium infection in patients and biological samples. The compounds
XX CC of the invention and antibodies directed against the Mycobacterium
XX CC proteins may be used in vaccines for immunisation against Mycobacterium
XX CC infections. The nucleic acids encoding the Mycobacterium proteins may be
XX CC used in gene therapy. AAU08159-AAU08188 represent peptides (ORF-2-1 to
XX CC ORF-2-30) which overlap to the open reading frame ORF-2
XX SQ Sequence 15 AA;
XX Query Match 2.8%; Score 6; DB 4; Length 15;
XX Best Local Similarity 100.0%; Pred.No. 1.7e+02;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 183 SDVLTA 188
XX DB |||||
XX 3 SDVLTA 8
XX RESULT 94
XX AAU08181
XX ID AAU08181 standard; peptide; 15 AA.
XX AC AAU08181;
XX AC AAU08181;
XX DT 17-DEC-2001 (first entry)
XX DE Mycobacterium tuberculosis overlapping peptide ORF-2-23.
XX XX Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX KW immunostimulant; mutant; mitein; ORF-2-23.
XX OS Mycobacterium tuberculosis.
XX OS Synthetic.
XX PN WO200162893-A2.
XX XX 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US005992.
XX XX 25-FEB-2000; 2000US-0185037P.
XX PR 08-AUG-2000; 2000US-0223828P.
XX XX (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Skeiky Y, Overdale P, Jen S, Lodes M;
XX WPI; 2001-536638/59.
XX An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
XX PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
XX PT response to and inhibiting development of a Mycobacterium infection.
XX PS Example 1; Page 134; 161pp; English.
XX CC The present invention relates to the isolation of Mycobacterium
XX CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
XX CC encoding them. The invention describes compounds and methods for the
XX CC diagnosis of tuberculosis or for inducing protective immunity against
XX CC tuberculosis. The compounds comprise at least one immunogenic portion of
XX CC one or more Mycobacterium proteins and nucleic acid molecules encoding
XX CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
XX CC encoding them can be used in diagnostic kits for the detection of
XX CC Mycobacterium infection in patients and biological samples. The compounds
XX CC of the invention and antibodies directed against the Mycobacterium
XX CC proteins may be used in vaccines for immunisation against Mycobacterium
XX CC infections. The nucleic acids encoding the Mycobacterium proteins may be
XX CC used in gene therapy. AAU08159-AAU08188 represent peptides (ORF-2-1 to
XX CC ORF-2-30) which overlap to the open reading frame ORF-2
XX SQ Sequence 15 AA;
XX Query Match 2.8%; Score 6; DB 4; Length 15;
XX Best Local Similarity 100.0%; Pred.No. 1.7e+02;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 183 SDVLTA 188
XX DB |||||
XX 3 SDVLTA 8
XX RESULT 94
XX AAU08181
XX ID AAU08181 standard; peptide; 15 AA.
XX AC AAU08181;
XX AC AAU08181;
XX DT 17-DEC-2001 (first entry)
XX DE Mycobacterium tuberculosis overlapping peptide ORF-2-23.
XX XX Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
XX KW immunostimulant; mutant; mitein; ORF-2-23.
XX OS Mycobacterium tuberculosis.
XX OS Synthetic.
XX PN WO200162893-A2.
XX XX 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US005992.
XX XX Use of laminin fragment and tissue plasminogen activator - for treating
XX XX and preventing blood clots and embolisms.

```


PS Claim 1; Page 15; 17pp; English.

XX The peptide is a laminin fragment comprising residues 2091-2108 of human

CC laminin with an additional Cys residue at the N-terminus. The peptide is

CC capable of activating tissue plasminogen activator and may be used in a

CC compn. with tPA for the prevention and treatment of blood clots and

CC embolisms. See also AAR25845-50. (Updated on 25-MAR-2003 to correct PN

CC field.)

XX Sequence 19 AA;

XX Query Match 2.8%; Score 6; DB 2; Length 19;

XX Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

XX Matches 6; Conservative 0; Mismatches 0;

OY 13 VAVSAD 18

Db 13 VAVSAD 18

RESULT 95

AAR44034

ID AAR44034 standard; peptide; 19 AA.

XX AC AAR44034;

XX 25-MAR-2003 (revised)

DT 03-DEC-1993 (first entry)

XX Laminin-derived peptide for Tc labelling, contains IKVAV.

XX integrin; cytoadhesiveness; mast cell; in vivo tumour localisation;

KW lung imaging; cell attachment; Technetium label; emphysema; cancer;

KW pulmonary disease; metal ion binding domain.

XX Synthetic.

XX Key Location/Qualifiers

FH Domain 11..15

FT /label= biological-function_domain

FT /note= "for lung imaging"

XX WO9312819-A1.

XX 08-JUL-1993.

XX 31-DEC-1992; 92WO-US011334.

XX 03-JAN-1992; 92US-00818476.

PR 03-JAN-1992; 92US-00818477.

PR 20-FEB-1992; 92US-00840077.

PR 30-DEC-1992; 92US-00998820.

PR 30-DEC-1992; 92US-00998910.

XX (RHOM-) RHOMED INC.

XX Rhodes BA, Zamora PO;

XX WPI; 1993-227063/28.

XX Metal labelled peptide(s) contg. binding and medically useful domains -

PT used in diagnosis and therapy of e.g. thrombus, cancer, infection,

PT inflammation, are also opt. combined to antibody.

XX Example 5; Page 33; 61pp; English.

XX This laminin-derived peptide contains the pentapeptide motif IKVAV, an

CC adhesive sequence from the laminin A-chain which is known to increase in

CC vitro adhesiveness of mast cells. The sequence is a preferred biological

CC function domain for peptides of the invention. In addition to a

CC biological function domain, the peptides contain a metal ion binding

CC domain (the N-terminal Cys residue in AAR44034) and are labelled with a

CC metal ion such as Technetium (esp. 99mTc). Peptides comprising IKVAV are

CC useful for lung imaging, e.g. to localise primary or metastatic cancerous

CC tumours (photon-abundant) or to monitor emphysema or fibrosis (photon-

CC deficient). (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 19 AA;

XX Query Match 2.8%; Score 6; DB 2; Length 19;

XX Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

XX Matches 6; Conservative 0; Mismatches 0;

OY 13 VAVSAD 18

Db 13 VAVSAD 18

RESULT 96

AAR92741

ID AAR92741 standard; peptide; 19 AA.

XX AC AAR92741;

XX 03-SEP-1996 (first entry)

DE IKVAV-containing sequence, for controlling cell distribution.

XX Control; distribution; bioartificial organ; BAO; cellular attachment;

KW neurotransmitter; hormone; cytokine; growth factor; enzyme.

XX Synthetic.

XX Key Location/Qualifiers

FH Peptide 11..15

FT /note= "Claimed core peptide"

XX WO9602646-A2.

XX 01-FEB-1996.

XX 20-JUL-1995; 95WO-US009281.

PR 20-JUL-1994; 94US-00279773.

PR 09-MAY-1995; 95US-00432898.

XX (CYTO-) CYTOTHERAPEUTICS INC.

XX Schinstine M, Shoichet MS, Gentile FT, Hamman JP, Holland LM;

PI Cain BM, Doherty EJ, Winn SR, Aebischer P;

XX WPI; 1996-105908/11.

XX Controlling distribution of cells in bio-artificial organs - e.g. by

PT treatment of cells, or growth surfaces, to inhibit proliferation, promote

PT differentiation or modulate adhesion, for in vivo prodn. of hormones,

PT neuro-transmitter(s) etc.

XX Claim 22; Page 71; 84pp; English.

XX The sequences given in AAR92739-41 are peptides which were used in the

CC method of the invention to control the distribution of cells within a

CC bioartificial organ (BAO). These peptides have been particularly useful

CC in promoting cellular attachment. These peptides are pref. bound to the

CC membrane of the BAO which is a biocompatible, permeable jacket. These

CC peptides act to control the distribution of the core of living cells

CC included in the BAO after in vivo implantation. BAO are used

CC therapeutically to produce e.g. neurotransmitters, hormones, cytokines,

CC growth factors, enzymes, etc

XX Sequence 19 AA;

XX Query Match 2.8%; Score 6; DB 2; Length 19;

XX Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

XX Matches 6; Conservative 0; Mismatches 0;

FH Key Location/Qualifiers
 FT Region 11.15
 XX /note= "Specifically claimed in Claim 22"
 XX US2001043923-A1.
 XX 22-NOV-2001.
 XX 07-MAR-2001; 2001US-00801237.
 XX 23-MAY-1995; 95US-00447997.
 XX (SCHI/) SCHINSTINE M.
 XX (SHOI/) SHOICHET M.S.
 XX (GENT/) GENTILE F.T.
 XX (HAMW/) HAMWANG J.P.
 XX (HOLL/) HOLLAND L.M.
 XX (CAIN/) CAIN B.M.
 XX (DOHE/) DOHERTY E.J.
 XX (WINN/) WINN S.R.
 XX (AEBI/) AEBISCHER P.
 XX Schinstine M, Shoichet MS, Gentile FT, Hamwang JP, Holland LM;
 XX Cain BM, Doherty EJ, Winn SR, Aebischer P;
 XX WPI; 2002-082383/11.
 XX Controlling distribution of cells within a bioartificial organ useful for
 XX delivering a biologically active molecule involves exposing the cells to
 XX a treatment that inhibits or promotes cell differentiation.
 XX Disclosure; Page 10; 27pp; English.
 XX The present invention relates to a method for controlling distribution of
 XX cells within a bioartificial organ. The method involves exposing the
 XX cells to a treatment that inhibits cell proliferation or promotes cell
 XX differentiation. The treatment is effective upon in vivo implantation in
 XX a host. The treatment of the method involves suspending the cells in a
 XX hydrogel matrix within the bioartificial organ, the matrix being
 XX derivatised with a proliferation-inhibiting or differentiation-inducing
 XX peptide. The present sequence is one such peptide, which also promotes
 XX cellular attachment. The present peptide recognises a cell surface
 XX receptor molecule involved in cell adhesion
 XX Sequence 19 AA;
 XX
 XX Query Match 2.8%; Score 6; DB 5; Length 19;
 XX Best Local Similarity 100.0%; Pred. No. 2e+02;
 XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 13 VAVSAD 18
 XX |||||
 XX Db 13 VAVSAD 18
 XX
 XX RESULT 100
 XX ABU90762
 XX ID ABU90762 standard; peptide; 19 AA.
 XX XX AC ABU90762;
 XX XX
 XX DT 11-JUL-2003 (first entry)
 XX XX
 XX Peptide #6 useful in cell-adhesive pattern for a biomedical device.
 XX Biomedical device; chemical surface pattern; substrate; prepatter;
 XX applied molecular self-assembly system; prefabricated pattern;
 XX cell-adhesive pattern; ligand; cell type; cell membrane receptor;
 XX analyte determination; biological sensing platform;
 XX bioanalytical sensing platform.
 XX Unidentified.
 XX OS
 XX

PN WO2003023401-A1.
 XX 20-MAR-2003.
 XX 12-SEP-2001; 2001WO-CH000548.
 XX 12-SEP-2001; 2001WO-CH000548.
 XX (EIDG-) EIDGENOESSISCHE TECH HOCHSCHULE ZUERICH.
 XX Textor M, Michel R, Voerces J, Hubbell JA, Lussi J;
 XX WPI; 2003-381450/36.
 XX Chemical surface patterns containing device for bioanalytical sensing
 XX platform, includes prepatterns of regions, and consecutively applied
 XX molecular self-assembly systems.
 XX Claim 73; Page 49; 69pp; English.
 XX The present invention relates to a device with chemical surface patterns
 XX with biochemical or biological relevance on substrates with prepatterns
 XX of at least two types of regions (alpha, beta), where at least two
 XX different, consecutively applied molecular self-assembly systems are used
 XX such that at least one of the applied assembly systems is specific to one
 XX type of the prefabricated patterns. The prefabricated patterns may
 XX comprise cell-adhesive patterns that contain specific ligands such as
 XX peptides, proteins and antibodies that are used to interact with specific
 XX cell type(s) or specific cell membrane receptor(s). The invention also
 XX discloses a method for the simultaneous qualitative or quantitative
 XX determination of analytes in sample(s) brought into contact with the
 XX measurement areas on a biological sensing platform. The device is useful
 XX for a bioanalytical sensing platform. ABU90757-ABU90875 represent
 XX peptides that may be used in cell-adhesive patterns for a device of the
 XX invention
 XX Sequence 19 AA;
 XX
 XX Query Match 2.8%; Score 6; DB 6; Length 19;
 XX Best Local Similarity 100.0%; Pred. No. 2e+02;
 XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 13 VAVSAD 18
 XX |||||
 XX Db 13 VAVSAD 18
 XX
 XX Search completed: August 6, 2004, 16:08:23
 XX Job time : 58 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2004, 16:07:23 ; Search time 18 seconds
(without alignments)
610.907 Million cell updates/sec

Title: US-10-024-955-7
Perfect score: 213
Sequence: 1 MKKFLIIAAVAVFVADPI.....VRKENTKVLAPAFKRELEKN 213

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	213	100.0	213	4	US-08-553-336A-7
2	25	11.7	215	3	US-08-462-778-2
3	25	11.7	215	4	US-08-553-336A-2
4	7	3.3	148	4	US-09-328-352-4312
5	7	3.3	204	1	US-08-292-945-8
6	7	3.3	204	1	US-08-252-073A-8
7	7	3.3	204	5	PCT-US93-1207A-8
8	7	3.3	293	4	US-09-252-991A-22215
9	7	3.3	381	4	US-08-540-236-3615
10	7	3.3	435	4	US-09-489-039A-10535
11	7	3.3	752	1	US-08-309-512-8
12	7	3.3	752	5	PCT-US92-08756A-8
13	6	2.8	10	1	US-08-029-333-31
14	6	2.8	10	4	US-08-391-820-31
15	6	2.8	12	1	US-07-616-910-54
16	6	2.8	12	1	US-08-405-200-2
17	6	2.8	12	1	US-08-405-200-15
18	6	2.8	12	5	PCT-US91-08497-54
19	6	2.8	15	4	US-09-073-009-73
20	6	2.8	15	4	US-09-073-009-75
21	6	2.8	15	4	US-09-073-010-73
22	6	2.8	15	4	US-09-073-010-75
23	6	2.8	18	1	US-07-895-252-10
24	6	2.8	18	1	US-08-072-283B-10
25	6	2.8	19	1	US-07-998-820-2
26	6	2.8	19	2	US-08-280-646-3
27	6	2.8	19	2	US-08-445-193-3

28	6	2.8	19	2	US-08-432-698-3	Sequence 3, Appli
29	6	2.8	19	2	US-08-447-810-3	Sequence 3, Appli
30	6	2.8	19	2	US-08-279-773-3	Sequence 3, Appli
31	6	2.8	19	2	US-08-747-137-56	Sequence 56, Appli
32	6	2.8	19	4	US-08-447-997-3	Sequence 3, Appli
33	6	2.8	19	4	US-08-801-237A-3	Sequence 3, Appli
34	6	2.8	19	5	PCT-US95-09282-3	Sequence 3, Appli
35	6	2.8	19	5	FCT-US95-09282-3	Sequence 3, Appli
36	6	2.8	57	4	US-08-562-737-124	Sequence 124, App
37	6	2.8	67	3	US-08-905-223-335	Sequence 335, App
38	6	2.8	67	4	US-09-107-532A-6800	Sequence 6800, Ap
39	6	2.8	71	4	US-05-543-681A-5866	Sequence 5866, Ap
40	6	2.8	88	4	US-03-634-238-234	Sequence 234, App
41	6	2.8	91	4	US-09-543-681A-6613	Sequence 6613, Ap
42	6	2.8	94	4	US-09-073-009-25	Sequence 25, Appl
43	6	2.8	94	4	US-09-073-009-27	Sequence 27, Appl
44	6	2.8	94	4	US-09-073-009-29	Sequence 29, Appl
45	6	2.8	94	4	US-09-073-010-25	Sequence 25, Appl
46	6	2.8	94	4	US-09-073-010-27	Sequence 27, Appl
47	6	2.8	94	4	US-09-073-010-29	Sequence 29, Appl
48	6	2.8	101	4	US-09-252-991A-24298	Sequence 24298, A
49	6	2.8	102	3	US-09-383-586-34	Sequence 34, Appl
50	6	2.8	113	4	US-09-540-236-2114	Sequence 2114, Ap
51	6	2.8	115	6	5164490-6	Patent No. 5164490
52	6	2.8	116	4	US-09-562-737-125	Sequence 125, App
53	6	2.8	128	4	US-03-199-637A-150	Sequence 150, App
54	6	2.8	130	2	US-08-944-449-2	Sequence 2, Appli
55	6	2.8	130	4	US-09-353-362-2	Sequence 2, Appli
56	6	2.8	133	2	US-08-891-848-15	Sequence 15, Appl
57	6	2.8	133	3	US-08-875-811-10	Sequence 10, Appl
58	6	2.8	134	2	US-08-891-848-14	Sequence 14, Appl
59	6	2.8	134	3	US-08-875-811-9	Sequence 9, Appli
60	6	2.8	134	4	US-09-446-959-5	Sequence 5, Appli
61	6	2.8	136	4	US-03-252-991A-25498	Sequence 25498, A
62	6	2.8	147	3	US-03-383-586-35	Sequence 35, Appl
63	6	2.8	152	4	US-09-446-959-7	Sequence 7, Appli
64	6	2.8	155	4	US-09-489-039A-11904	Sequence 11904, A
65	6	2.8	159	4	US-09-107-532A-4926	Sequence 4926, Ap
66	6	2.8	163	4	US-09-621-976-4550	Sequence 4550, Ap
67	6	2.8	164	4	US-09-489-039A-11750	Sequence 11750, A
68	6	2.8	168	4	US-03-198-452A-1213	Sequence 1213, Ap
69	6	2.8	170	1	US-08-460-739-3	Sequence 3, Appli
70	6	2.8	170	2	US-08-483-101-11	Sequence 11, Appl
71	6	2.8	178	4	US-09-446-959-10	Sequence 10, Appl
72	6	2.8	179	4	US-09-107-532A-3794	Sequence 3794, Ap
73	6	2.8	180	4	US-09-543-681A-7669	Sequence 7669, Ap
74	6	2.8	183	4	US-05-543-681A-5384	Sequence 5384, Ap
75	6	2.8	186	3	US-09-230-637-96	Sequence 36, Appl
76	6	2.8	186	4	US-09-134-000C-5358	Sequence 5358, Ap
77	6	2.8	186	4	US-09-142-530C-7	Sequence 7, Appli
78	6	2.8	186	4	US-09-138-452A-444	Sequence 444, App
79	6	2.8	195	4	US-09-134-001C-4319	Sequence 4319, Ap
80	6	2.8	201	4	US-08-529-055-21	Sequence 21, Appl
81	6	2.8	204	4	US-09-134-000C-4079	Sequence 4079, Ap
82	6	2.8	205	4	US-03-418-710-63	Sequence 63, Appl
83	6	2.8	211	4	US-08-861-774E-44	Sequence 44, Appl
84	6	2.8	215	4	US-09-143-127-7	Sequence 7, Appli
85	6	2.8	222	1	US-08-143-127-1	Sequence 1, Appli
86	6	2.8	226	1	US-08-028-463-4	Sequence 4, Appli
87	6	2.8	226	1	US-08-461-836-4	Sequence 4, Appli
88	6	2.8	230	4	US-09-489-039A-10204	Sequence 10204, A
89	6	2.8	231	4	US-03-495-406-35	Sequence 35, Appl
90	6	2.8	233	2	US-08-557-309B-40	Sequence 40, Appl
91	6	2.8	233	3	US-08-834-306-40	Sequence 40, Appl
92	6	2.8	233	3	US-08-993-674A-40	Sequence 40, Appl
93	6	2.8	233	4	US-09-256-976-40	Sequence 8, Appli
94	6	2.8	235	4	US-09-374-827-8	Sequence 28804, A
95	6	2.8	236	4	US-09-252-991A-28804	Sequence 28804, A
96	6	2.8	237	4	US-09-252-991A-29211	Sequence 29211, A
97	6	2.8	240	4	US-03-252-991A-28714	Sequence 28714, A
98	6	2.8	252	4	US-09-522-714-20	Sequence 20, Appl
99	6	2.8	252	4	US-09-489-039A-13690	Sequence 13690, A
100	6	2.8	256	4	US-09-543-681A-7108	Sequence 7108, Ap

ALIGNMENTS

RESULT 1

US-08-553-336A-7
; Sequence 7, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:

APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,540

FILING DATE: 22-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jane E. Remillard

REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: IMI-032CF2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-553-336A-7

Query Match 100.0%; Score 213; DB 4; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.1e-207;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKFKLLIAAVFAVAVSADPHYDKITEINKAIDDAIAAEQSEITDPMKVPDPAKFER	60
DB	1	MKFKLLIAAVFAVAVSADPHYDKITEINKAIDDAIAAEQSEITDPMKVPDPAKFER	60
QY	61	HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEEGIVKHAHLIGVHDDIVSMEYDLAY	120
DB	61	HVGIVDFKGLAMRNIEARGLKQMKRQGDANVKGEEGIVKHAHLIGVHDDIVSMEYDLAY	120
QY	121	KLGLDHPHTTHVISIQDFVVALSLEISDEGNITMTSFEVRQFANVNVNHIGLSILDPIFG	180
DB	121	KLGLDHPHTTHVISIQDFVVALSLEISDEGNITMTSFEVRQFANVNVNHIGLSILDPIFG	180
QY	181	VLSDVLTAIFQDITVRKEMTKVLAPAFKRELEKN	213
DB	181	VLSDVLTAIFQDITVRKEMTKVLAPAFKRELEKN	213

RESULT 2

US-08-462-778-2
; Sequence 2, Application US/08462778
; Patent No. 6077517

GENERAL INFORMATION:

APPLICANT: Thomas, Wayne R.
APPLICANT: Chua, Kaw-Yan
TITLE OF INVENTION: Allergenic Protein and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,778
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/031,141
FILING DATE: 12 March 1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras

REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-053CP (IMI-032CP)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 215 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-462-778-2

Query Match 11.7%; Score 25; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	155	TSFEVRQFANVNVNHIGLSILDPIF	179
DB	155	TSFEVRQFANVNVNHIGLSILDPIF	179

RESULT 3

US-08-553-336A-2
; Sequence 2, Application US/08553336A
; Patent No. 6413738
; GENERAL INFORMATION:

APPLICANT: Wayne R. Thomas and Kaw-Yan Chua
TITLE OF INVENTION: Allergenic Proteins and Peptides From
; TITLE OF INVENTION: House Dust Mite and Uses Therefor
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,336A
FILING DATE: 10-JUN-1996

```
;
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,540
; FILING DATE: 22-JUNE-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: IMI-032CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-053-336A-2

Query Match 11.7%; Score 25; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.3e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TSFEVQFANVNVHIGLSILDPF 179
Db 155 TSFEVQFANVNVHIGLSILDPF 179

RESULT 4
US-09-328-352-4312
; Sequence 4312, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4312
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4312

Query Match 3.3%; Score 7; DB 4; Length 148;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGGLSIL 175
Db 21 IGGLSIL 27

RESULT 5
US-08-292-945-8
; Sequence 8, Application US/08292945
; Patent No. 5585478
; GENERAL INFORMATION:
; APPLICANT: Lim, Bing
; APPLICANT: Lelias, Jean-Michel
; APPLICANT: Adra, Chaker
; APPLICANT: Ko, Jong
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30309-4530

Query Match 3.3%; Score 7; DB 4; Length 148;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 IGGLSIL 175
Db 21 IGGLSIL 27

RESULT 6
US-08-252-073A-8
; Sequence 8, Application US/08252073A
; Patent No. 5767073
; GENERAL INFORMATION:
; APPLICANT: Bing Lim, et al.
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center,
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,073A
; FILING DATE: 1-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/292,945
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: December 10, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MED107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
; US-08-292-945-8

Query Match 3.3%; Score 7; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 6
US-08-252-073A-8
; Sequence 8, Application US/08252073A
; Patent No. 5767073
; GENERAL INFORMATION:
; APPLICANT: Bing Lim, et al.
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center,
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,073A
; FILING DATE: 1-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
```

; REFERENCE/DOCKET NUMBER: MED107cip
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)873-8794
; TELEFAX: (404)873-8795
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
; US-08-252-073A-8

Query Match 3.3%; Score 7; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 7
PCT-US93-12074-8
; Sequence 8, Application PC/TUS9312074
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12074
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,337
; FILING DATE: 10-DEC-1992
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bovine
; TISSUE TYPE: Brain
; PCT-US93-12074-8

Query Match 3.3%; Score 7; DB 5; Length 204;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSADP 19
Db 59 VAVSADP 65

RESULT 8
US-09-252-991A-22215
; Sequence 22215, Application US/09252991A

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22215
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-22215

Query Match 3.3%; Score 7; DB 4; Length 293;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AFAVSA 17
Db 110 AFAVSA 116

RESULT 9
US-09-540-236-3615
; Sequence 3615, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3615
; LENGTH: 381
; TYPE: PRT
; ORGANISM: M.catarrhalis
; US-09-540-236-3615

Query Match 3.3%; Score 7; DB 4; Length 381;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 PTHVIS 133
Db 82 PTHVIS 88

RESULT 10
US-09-489-039A-10535
; Sequence 10535, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10535
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-10535

Query Match 3.3%; Score 7; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAVAFVA 14
Db 363 AAVAFVA 369

RESULT 11

US-08-309-512-8
; Sequence 8, Application US/08309512
; Patent No. 5759828
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Benzman, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoun, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94025

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,512
; FILING DATE:
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: 8145-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694

INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum

US-08-309-512-8

Query Match 3.3%; Score 7; DB 1; Length 752;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEQSET 45
Db 304 AIEQSET 310

RESULT 12

PCT-US92-08756A-8

; Sequence 8, Application PC/TUS9208756A
; GENERAL INFORMATION:
; APPLICANT: Tal, Ronny
; APPLICANT: Benzman, Moshe
; APPLICANT: Gelfand, David H.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoun, Roger D.
; APPLICANT: Wong, Hing C.
; TITLE OF INVENTION: CYCLIC DIGUANYLATE
; TITLE OF INVENTION: METABOLIC ENZYMES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Limbach and Limbach
; STREET: 2001 Perry Building
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0,
; SOFTWARE: Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08756A
; FILING DATE: 19921014
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,218
; FILING DATE: 29-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: WEYR 20050 USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-433-4150
; TELEFAX: 415-433-8716
; TELEX: 278356

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter xylinum

PCT-US92-08756A-8

Query Match 3.3%; Score 7; DB 5; Length 752;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 AIEQSET 45
Db 304 AIEQSET 310

RESULT 13

US-08-029-333-31
; Sequence 31, Application US/08029333
; Patent No. 5399667
; GENERAL INFORMATION:
; APPLICANT: Frazier, William A.
; APPLICANT: Kosfeld, Minh D.
; TITLE OF INVENTION: Thrombospondin Receptor Binding Peptides
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SG
; STREET: 800 N. Lindbergh Blvd.

;; CITY: St. Louis
;; STATE: Missouri
;; COUNTRY: USA
;; ZIP: 63167
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/029,333
;; FILING DATE: 19930305
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyer, Scott J.
;; REGISTRATION NUMBER: 25,275
;; REFERENCE/DOCKET NUMBER: 07-24(982)A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (314)694-3117
;; TELEFAX: (314)694-5435
;; INFORMATION FOR SEQ ID NO: 31:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-029-333-31

Query Match 2.8%; Score 6; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VAVSAD 18
Db 4 VAVSAD 9

RESULT 14
US-08-391-820-31
; Sequence 31, Application US/08391820
; Patent No. 6469138
; GENERAL INFORMATION:
; APPLICANT: Frazier, William A.
; TITLE OF INVENTION: Thrombospondin Receptor Binding Peptides
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SG
; STREET: 800 N. Lindbergh Blvd.
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63167
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,820
; FILING DATE: 21-Feb-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/029,333
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyer, Scott J.
; REGISTRATION NUMBER: 25,275
; REFERENCE/DOCKET NUMBER: 07-24(982)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)694-3117
; TELEFAX: (314)694-5435
; INFORMATION FOR SEQ ID NO: 31:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 31:
;; US-08-391-820-31

Query Match 2.8%; Score 6; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 VAVSAD 18
Db 4 VAVSAD 9

RESULT 15
US-07-616-910-54
; Sequence 54, Application US/07616910
; Patent No. 5223484
; GENERAL INFORMATION:
; APPLICANT: Fleming, Patrick J.
; TITLE OF INVENTION: Peptide Which Regulates Weight Gain in Mammals
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/616,910
; FILING DATE: 19901121
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5223484man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-069-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-616-910-54

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 YDLAYK 121
Db 2 YDLAYK 7

RESULT 16
US-08-405-200-2
; Sequence 2, Application US/08405200
; Patent No. 5696229
; GENERAL INFORMATION:

APPLICANT: LAURIE, Gordon W
APPLICANT: MATTER, Michelle L
APPLICANT: CHEN, Lanlin
TITLE OF INVENTION: POLYPEPTIDE WITH LAMININ CELL ADHESION
TITLE OF INVENTION: AND MORPHOGENESIS ACTIVITY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,200
FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 494-150-0
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-200-2

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 6 VAVSAD 11

RESULT 17
US-08-405-200-15
Sequence 15, Application US/08405200
Patent No. 5696229
GENERAL INFORMATION:
APPLICANT: LAURIE, Gordon W
APPLICANT: MATTER, Michelle L
APPLICANT: CHEN, Lanlin
TITLE OF INVENTION: POLYPEPTIDE WITH LAMININ CELL ADHESION
TITLE OF INVENTION: AND MORPHOGENESIS ACTIVITY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,200

FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 494-150-0
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-200-15

Query Match 2.8%; Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 VAVSAD 18
Db 6 VAVSAD 11

RESULT 18
PCT-US91-08497-54
Sequence 54, Application PC/TUS9108497
GENERAL INFORMATION:
APPLICANT: Fleming, Patrick J.
APPLICANT: Kent, Ute M.
TITLE OF INVENTION: Peptide Which Regulates Weight Gain in Mammals
TITLE OF INVENTION: Mammals
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08497
FILING DATE: 19911121
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/616,910
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, Norman F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-069-0 PCT
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US91-08497-54

Query Match 2.8%; Score 6; DB 5; Length 12;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YDLAYK 121
DB 2 YDLAYK 7

RESULT 19

US-09-073-009-73
; Sequence 73, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.009
FILING DATE: 05-MAY-1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis

US-09-073-009-73

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 8 SDVLTA 13

RESULT 20

US-09-073-009-75
; Sequence 75, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144

US-09-073-009-73

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 8 SDVLTA 13

RESULT 21

US-09-073-010-73
; Sequence 73, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.010
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392

TUBERCULOSIS AND ME

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-882-6031

INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis

US-09-073-009-75

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 3 SDVLTA 8

RESULT 21

US-09-073-010-73
; Sequence 73, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.010
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392

US-09-073-009-75

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 3 SDVLTA 8

RESULT 21

US-09-073-010-73
; Sequence 73, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.010
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392

US-09-073-009-75

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 3 SDVLTA 8

RESULT 21

US-09-073-010-73
; Sequence 73, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.010
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392

US-09-073-009-75

Query Match 2.8%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 SDVLTA 188
DB 3 SDVLTA 8

RESULT 21